PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12Q 1/68, C12P 21/06, C12N 15/12, A61K 38/17, C07K 14/47, 14/705

A1

(11) International Publication Number: **WO 98/23780**

(43) International Publication Date: 4 June 1998 (04.06.98)

(21) International Application Number: PCT/US97/21614

(22) International Filing Date: 25 November 1997 (25.11.97)

(30) Priority Data:
60/032,043 26 November 1996 (26.11.96) US

(71) Applicant: JOSLIN DIABETES CENTER, INC. [US/US]; One Joslin Place, Boston, MA 02215 (US).

(72) Inventor: KROLEWSKI, Andrzej, S.; 639 Great Plain Avenue, Needham, MA 02192 (US).

(74) Agent: GREER, Helen; Banner & Witcoff, Ltd., One Financial Center, Boston, MA 02111 (US).

(81) Designated States: CA, JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: METHODS FOR DIAGNOSING AND TREATING DIABETES

(57) Abstract

A method for determining if an animal is at risk for diabetes is described. An animal is provided and an aspect of HNF4 metabolism or structure is evaluated in the animal. An abnormality in the aspect of HNF4 metabolism or structure is diagnostic of being at risk for diabetes. Also described are methods for evaluating an agent for use in treating diabetes, methods for treating diabetes, and methods for treating a cell having an abnormality in structure or metabolism of HNF4. Pharmaceutical compositions and vaccine compositions are also provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

METHODS FOR DIAGNOSING AND TREATING DIABETES

This application claims the benefit of U.S. Provisional Application No. 60/032,043 filed November 26, 1996.

The U.S. Government has a paid-up license in this invention and the right in limited circumstances to require the patent owner to license others on reasonable terms as provided for by the terms of Grant No. R01-DK47475 awarded by the National Institutes of Health.

Field of the Invention

This invention relates generally to treatments, diagnoses, and therapeutic agents for diabetes.

Background of the Invention

Diabetes is a serious metabolic disorder. Insulin-dependent diabetes, known as Type I diabetes represents about 5-10% of all patients with diabetes. Non-insulin-dependent diabetes (NIDDM), known as Type II diabetes, is the most common of all metabolic disorders. It has been reported that NIDDM currently affects about 6-7% of the U.S. population, with a cumulative risk of about 17% by age 80. The metabolic derangements created by chronic hyperglycemia, plus the strong association between NIDDM, obesity, hypertension and hyperlipidemia, result in extensive long-term complications, including a high rate of cardiovascular death and amputation due to accelerated atherosclerosis, as well as the typical complications of retinopathy, nephropathy and neuropathy. The economic burden created by diabetes and its complications has been estimated at over $120 billion annually in the United States alone.

Genetic susceptibility plays an important role in the development of NIDDM. Rare forms of NIDDM, known as maturity-onset diabetes of the young (MODY), (which have been estimated to account for about 1-3% of NIDDM), generally develop before age 30. Three different MODY genes have been partially mapped: MODY 1 is reported to be linked to chromosome 20q (to an as yet unidentified gene), MODY 2 is reported to be linked to the glucokinase gene on chromosome 7p, and MODY 3 is reported to be linked to chromosome 12q. The common forms of NIDDM (which have been estimated to account for about 95% of individuals with diabetes), generally occur in middle or old age. The gene(s) responsible for the common form of NIDDM has not been previously identified.

Summary of the Invention

It is an object of the invention to provide a method for determining if an animal is at risk for diabetes, in particular NIDDM.

It is yet another object of the invention to provide a method for evaluating an agent for use in treating diabetes.

It is yet another object of the invention to provide a method for treating diabetes.

Still another object of the invention is to utilize the HNF4 gene and/or polypeptide, and fragments, analogs and variants thereof, to aid in the treatment, diagnosis and/or identification of therapeutic agents for diabetes.

In one aspect, the invention features a method for determining if an animal is at risk for diabetes. An animal is provided. An aspect of HNF4 metabolism or structure is evaluated in the animal. An abnormality in the aspect of HNF4 metabolism or structure is diagnostic of being at risk for diabetes.

Another aspect of the invention is a method for evaluating an agent for use in treating diabetes. A test cell, cell-free system or animal, having a non-wild type pattern of HNF4 metabolism is provided. An agent is provided. The agent is administered to the test cell, cell-free system or animal in a therapeutically effective amount. The effect of the agent on an aspect of HNF4 metabolism is evaluated. A change in the aspect of HNF4 metabolism is indicative of the usefulness of the agent in treating diabetes.

Another aspect of the invention is a method for evaluating an agent for the ability to alter the binding of HNF4 polypeptide to a binding molecule. An agent is provided. An HNF4 polypeptide is provided. A binding molecule is provided. The agent, HNF4 polypeptide and binding molecule are combined. The formation of a complex comprising the HNF4 polypeptide and binding molecule is detected. An alteration in the formation of the complex in the presence of the agent as compared to in the absence of the agent is indicative of the agent altering the binding of the HNF4 polypeptide to the binding molecule.

Another aspect of the invention is a method for evaluating an agent for the ability to bind to an HNF4 polypeptide. An agent is provided. An HNF4 polypeptide is provided. The agent is contacted with the HNF4 polypeptide. The ability of the agent to bind to the HNF4 polypeptide is evaluated.

Another aspect of the invention is a method for evaluating an agent for the ability to bind to a nucleic acid encoding an HNF4 regulatory sequence. An agent is provided. A nucleic acid

encoding an HNF4 regulatory sequence is provided. The agent is contacted with the nucleic acid. The ability of the agent to bind to the nucleic acid is evaluated.

Another aspect of the invention is a method for treating diabetes in an animal. An animal in need of treatment for diabetes is provided. An agent capable of altering an aspect of HNF4 structure or metabolism is provided. The agent is administered to the animal in a therapeutically effective amount such that treatment of the diabetes occurs.

Another aspect of the invention is a method for treating an animal at risk for diabetes. An animal at risk for diabetes is provided. An agent capable of altering an aspect of HNF4 structure or metabolism is provided. The agent is administered to the animal in a therapeutically effective amount such that treatment of the animal occurs.

Another aspect of the invention is a method for treating a cell having an abnormality in structure or metabolism of HNF4. A cell having an abnormality in structure or metabolism of HNF4 is provided. An agent capable of altering an aspect of HNF4 structure or metabolism is provided. The agent is administered to the cell in a therapeutically effective amount such that treatment of the cell occurs.

Another aspect of the invention is a pharmaceutical composition for treating diabetes in an animal comprising a therapeutically effective amount of an agent, the agent being capable of altering an aspect of HNF4 metabolism or structure in the animal so as to result in treatment of the diabetes, and a pharmaceutically acceptable carrier.

Another aspect of the invention is a vaccine composition for treating diabetes in an animal comprising a therapeutically effective amount of an agent, the agent being capable of altering an aspect of HNF4 metabolism or structure in the animal so as to result in treatment of the diabetes, and a pharmaceutically acceptable carrier.

Another aspect of the invention is a method of making a fragment or analog of HNF4 polypeptide, the fragment or analog having a transcription activating activity of a naturally occurring HNF4 polypeptide for a gene involved in insulin metabolism. An HNF4 polypeptide is provided. The sequence of the HNF4 polypeptide is altered. The altered HNF4 polypeptide is tested for the transcription activating activity.

Yet another aspect of the invention is a method of making an HNF4 polypeptide or fragment or analog thereof having a non-wild type activity as a transcription activator for a gene involved in insulin metabolism. The sequence of an HNF4 polypeptide is altered. The altered HNF4 polypeptide is tested for the non-wild type activity.

The above and other features, objects and advantages of the present invention will be better understood by a reading of the following specification in conjunction with the drawings.

Brief Description of the Drawings

Fig. 1 depicts the nucleotide sequence encoding human HNF4α1.

Fig. 2 depicts the nucleotide sequence encoding human HNF4α2.

Fig. 3 depicts the nucleotide sequence encoding human HNF4α4.

Fig. 4 depicts the nucleotide sequence encoding human HNF4γ.

Fig. 5 depicts a multipoint nonparametric linkage analysis of NIDDM with markers in the MODY 1 region of human chromosome 20q.

Detailed Description

This invention provides a method for determining if an animal is at risk for diabetes. An animal is provided. An aspect of HNF4 metabolism or structure is evaluated in the animal. An abnormality in the aspect of HNF4 metabolism or structure is diagnostic of being at risk for diabetes.

By diabetes is meant a disease or condition which is an insulin-related disorder in which the body does not metabolize glucose properly. Type I diabetes, known as insulin-dependent diabetes, can result from a lack of insulin secreting beta cells. Type II diabetes, known as non-insulin-dependent diabetes (NIDDM), can result from, e.g., resistance to insulin or insufficient production of insulin, or insufficient secretion of insulin. NIDDM has a variety of forms which are grouped as common forms and rare forms. In preferred embodiments of the invention, diabetes is any of the forms of NIDDM. In other preferred embodiments, diabetes is any of the common forms of NIDDM. In other preferred embodiments, diabetes is any of the rare forms of NIDDM referred to as maturity-onset diabetes of the young (MODY), preferably MODY 1.

By animal is meant human as well as non-human animals. Non-human animals include, e.g., mammals, birds, reptiles, amphibians and fish. Preferably, the non-human animal is a mammal, e.g., a rodent, e.g., a mouse or rat, a rabbit, a monkey, or a pig. An animal also includes transgenic non-human animals. The term transgenic animal is meant to include an animal that has gained new genetic information from the introduction of foreign DNA, i.e., partly or entirely heterologous DNA, into the DNA of its cells; or introduction of a lesion, e.g., an in vitro induced mutation, e.g., a deletion or other chromosomal rearrangement into the DNA of its

cells; or introduction of homologous DNA into the DNA of its cells in such a way as to alter the genome of the cell into which the DNA is inserted, e.g., it is inserted at a location which differs from that of the natural gene or its insertion results in a knockout. The animal may include a transgene in all of its cells including germ line cells, or in only one or some of its cells. Transgenic animals of the invention can serve as a model for studying diabetes.

In certain embodiments, the determination for being at risk for diabetes is done in a prenatal animal.

HNF4, hepatocyte nuclear factor 4, is one of several positive-acting transcription factors that has been reported to play a role in regulating the transcription of, e.g., many of the genes essential to the functioning of the liver. (Reviewed in Sladek, F.M., in Liver Gene Expression, eds. Tronche and Yaniv, R.G. Landes Co., Austin, TX, pp. 207-230 (1994)). Based on sequence and structural similarity, HNF4 has been classified as a member of the nuclear receptor superfamily, which contains ligand-dependent transcription factors, i.e., they bind to DNA and to a ligand. The ligand for HNF4 has not yet been identified, and therefore HNF4 is referred to as an orphan receptor. Isoforms of HNF4 exist which are reported to be derived from alternative splicing events. Three isoforms have been mapped to human chromosome 20, HNF4α1, HNF4α2 and HNF4α4, and one isoform to human chromosome 8, HNF4γ (Drewes et al., Mol. and Cell. Biol. 16(3):925-931 (1996); Kritis et al., Gene 173:275-280 (1996)). The sequences of these human HNF4 isoforms have been reported (Drewes et al., Molec. and Cell Biol. 16(3):925-931 (1996); Kritis et al., Gene 173:275-280 (1996)). The nucleic acid sequences of these isoforms, HNF4α1, HNF4α2, HNF4α4 and HNFγ, are shown in Figs. 1-4, respectively (SEQ. ID NO:1, SEQ. ID NO:2, SEQ. ID NO:3 and SEQ. ID NO:4). In preferred embodiments, the aspect of HNF4 metabolism or structure that is evaluated is related to HNF4 from chromosome 20.

By HNF4 metabolism is meant any aspect of the production, release, expression, function, action, interaction or regulation of HNF4. The metabolism of HNF4 includes modifications, e.g., covalent or non-covalent modifications, of HNF4 polypeptide. The terms peptides, proteins and polypeptides are used interchangeably herein. The metabolism of HNF4 includes modifications, e.g., covalent or non-covalent modifications, that HNF4 induces in other substances. The metabolism of HNF4 also includes changes in the distribution of HNF4 polypeptide, and changes HNF4 induces in the distribution of other substances.

Any aspect of HNF4 metabolism can be evaluated. The methods used are standard

techniques known to those skilled in the art and can be found in standard references, e.g., Ausubel et al., ed., Current Protocols in Mol. Biology, New York: John Wiley & Sons, 1990. (Drewes et al., Mol. and Cell. Biol. 16(3):925-931 (1996)). Preferred examples of HNF4 metabolism that can be evaluated include the binding activity of HNF4 polypeptide to a binding molecule; the transactivation activity of HNF4 polypeptide on a target gene; the level of HNF4 protein; the level of HNF4 mRNA; or the level of HNF4 phosphorylation. By binding molecule is meant any molecule to which HNF4 can bind, e.g., a nucleic acid, e.g., a DNA regulatory region, a protein, a metabolite, a peptide mimetic, a non-peptide mimetic, an antibody, or any other type of ligand. Binding can be shown, e.g., by electrophoretic mobility shift analysis (EMSA). Transactivation of a target gene by HNF4 can be determined, e.g., in a transient transfection assay in which the promoter of the target gene is linked to a reporter gene, e.g., β-galactosidase or luciferase, and co-transfected with an HNF4 expression vector. Examples of target genes are HNF1, apolipoproteins A1, AII, AIV, B and CIII. Other target genes are reported in Sladek, F.M., in Liver Gene Expression, eds. Tronche and Yaniv, R.G. Landes Co., Austin, TX, pp. 207-230 (1994). Such evaluations can be done in vitro or in vivo. Levels of HNF4 protein, mRNA or phosphorylation, can, e.g., be measured in a sample, e.g., a tissue sample, e.g., in a classic insulin sensitive tissue such as muscle, fat or liver tissue, or in a more easily accessible tissue such as circulating blood cells or fibroblasts.

In certain embodiments, an aspect of HNF4 structure is evaluated, e.g., HNF4 gene structure or HNF4 protein structure. For example, primary, secondary or tertiary structures can be evaluated. For example, the DNA sequence of the gene is determined and/or the amino acid sequence of the protein is determined. Standard cloning and sequencing methods can be used as are known to those skilled in the art. In certain embodiments, the binding activity of an antisense nucleic acid with the cellular HNF4 mRNA and/or genomic DNA is determined using standard methods known to those skilled in the art so as to detect the presence or absence of the target mRNA or DNA sequences to which the antisense nucleic acid would normally specifically bind.

The invention also includes a method for evaluating an agent for use in treating diabetes. A test cell, cell-free system or animal having a non-wild type pattern of HNF4 metabolism is provided. An agent is provided. The agent is administered to the test cell, cell-free system or animal in a therapeutically effective amount. The effect of the agent on an aspect of HNF4 metabolism is evaluated. A change in the aspect of HNF4 metabolism is indicative of the usefulness of the agent in treating diabetes. In certain other embodiments, the method uses test

cell, cell-free system or animal having a wild-type pattern of HNF4 metabolism.

By cell is meant a cell or a group of cells, or a cell that is part of an animal. The cell can be a human or non-human cell. Cell is also meant to include a transgenic cell. The cell can be obtained, e.g., from a culture or from an animal. Animals are meant to include, e.g., natural animals and non-human transgenic animals. In certain embodiments, the transgenic cell or non-human transgenic animal has an HNF4 transgene, or fragment or analog thereof. In certain embodiments, the transgenic cell or non-human transgenic animal has a knockout for the HNF4 gene.

A non-wild type pattern of HNF4 metabolism can result, e.g., from under-expression, over-expression, no expression, or a temporal, site or distribution change. Such a non-wild type pattern can result, e.g., from one or more mutations in the HNF4 gene, in a binding molecule gene, or in any other gene which directly or indirectly affects HNF4 metabolism. A mutation is meant to include, e.g., an alteration, e.g., in gross or fine structure, in a nucleic acid. Examples include single base pair alterations, e.g., missense or nonsense mutations, frameshifts, deletions, insertions and translocations. Mutations can be dominant or recessive. Mutations can be homozygous or heterozygous.

An agent is meant to include, e.g., any substance, e.g., an anti-diabetic drug. The agent of this invention preferably can change an aspect of HNF4 metabolism. Such change can be the result of any of a variety of events, including, e.g., preventing or reducing interaction between HNF4 and a binding molecule; inactivating HNF4 and/or the binding molecule, e.g., by cleavage or other modification; altering the affinity of HNF4 and the binding molecule for each other; diluting out HNF4 and/or the binding molecule; preventing expression of HNF4 and/or the binding molecule; reducing synthesis of HNF4 and/or the binding molecule; synthesizing an abnormal HNF4 and/or binding molecule; synthesizing an alternatively spliced HNF4 and/or binding molecule; preventing or reducing proper conformational folding of HNF4 and/or the binding molecule; modulating the binding properties of HNF4 and/or the binding molecule; interfering with signals that are required to activate or deactivate HNF4 and/or the binding molecule; activating or deactivating HNF4 and/or the binding molecule at the wrong time; or interfering with other receptors, ligands or other molecules which are required for the normal synthesis or functioning of HNF4 and/or the binding molecule.

Examples of agents include HNF4 polypeptide or a biologically active fragment or analog thereof; a nucleic acid encoding HNF4 polypeptide or a biologically active fragment

thereof; a nucleic acid comprising a nucleotide sequence as set forth in Fig. 1 (SEQ. ID NO:1), or a biologically active fragment thereof; a nucleic acid encoding an HNF4 regulatory sequence or a biologically active fragment thereof; a binding molecule for HNF4 polypeptide; a binding molecule for HNF4 nucleic acid, the HNF4 nucleic acid being, e.g., a nucleic acid comprising a regulatory region for HNF4 or a nucleic acid comprising a structural region for HNF4 or a biologically active fragment of HNF4; an antisense nucleic acid; a mimetic of HNF4 or a binding molecule; an antibody for HNF4 or a binding molecule; a metabolite; or an inhibitory carbohydrate or glycoprotein. In certain embodiments, the agent is an antagonist, agonist or super agonist.

HNF4 allows a search for natural or artificial ligands to regulate insulin in the treatment of diabetics. In certain embodiments, the agent is a natural ligand for HNF4. In certain embodiments, the agent is an artificial ligand for HNF4.

By analog is meant a compound that differs from naturally occurring HNF4 in amino acid sequence or in ways that do not involve sequence, or both. Analogs of the invention generally exhibit at least about 90% homology, preferably at least about 95% homology, and most preferably at least about 99% homology, with a segment of 20 amino acid residues, preferably with more than 40 amino acid residues, or more preferably yet with substantially the entire sequence of a naturally occurring HNF4 sequence. Non-sequence modifications include, e.g., in vivo or in vitro chemical derivatizations of HNF4. Non-sequence modifications include, e.g., changes in phosphorylation, acetylation, methylation, carboxylation, or glycosylation. Methods for making such modifications are known to those skilled in the art. For example, phosphorylation can be modified by exposing HNF4 to phosphorylation-altering enzymes, e.g., kinases or phosphatases.

Preferred analogs include HNF4 or biologically active fragments thereof, whose sequences differ from the wild-type sequence by one or more conservative amino acid substitutions or by one or more non-conservative amino acid substitutions, deletions, or insertions which do not abolish HNF4 biological activity. Conservative substitutions typically include the substitution of one amino acid for another with similar characteristics, e.g., substitutions within the following groups: valine, glycine; glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. Other conservative substitutions are shown in Table 1.

Table 1

CONSERVATIVE AMINO ACID SUBSTITUTIONS

For Amino Acid	Code	Replace with any of
Alanine	A	D-Ala, Gly, beta-Ala, L-Cys, D-Cys
Arginine	R	D-Arg, Lys, D-Lys, homo-Arg, D-homo-Arg, Met, Ile, D-Met, D-Ile, Orn, D-Orn
Asparagine	N	D-Asn, Asp, D-Asp, Glu, D-Glu, Gln, D-Gln
Aspartic Acid	D	D-Asp, D-Asn, Asn, Glu, D-Glu, Gln, D-Gln
Cysteine	C	D-Cys, S-Me-Cys, Met, D-Met, Thr, D-Thr
Glutamine	Q	D-Gln, Asn, D-Asn, Glu, D-Glu, Asp, D-Asp
Glutamic Acid	E	D-Glu, D-Asp, Asp, Asn, D-Asn, Gln, D-Gln
Glycine	G	Ala, D-Ala, Pro, D-Pro, β-Ala Acp
Isoleucine	I	D-Ile, Val, D-Val, Leu, D-Leu, Met, D-Met
Leucine	L	D-Leu, Val, D-Val, Leu, D-Leu, Met, D-Met
Lysine	K	D-Lys, Arg, D-Arg, homo-Arg, D-homo-Arg, Met, D-Met, Ile, D-Ile, Orn, D-Orn
Methionine	M	D-Met, S-Me-Cys, Ile, D-Ile, Leu, D-Leu, Val, D-Val
Phenylalanine	F	D-Phe, Tyr, D-Thr, L-Dopa, His, D-His, Trp, D-Trp, Trans-3,4, or 5-phenylproline, cis-3,4, or 5-phenylproline
Proline	P	D-Pro, L-I-thioazolidine-4-carboxylic acid, D-or L-1-oxazolidine-4-carboxylic acid
Serine	S	D-Ser, Thr, D-Thr, allo-Thr, Met, D-Met, Met(O), D-Met(O), L-Cys, D-Cys
Threonine	T	D-Thr, Ser, D-Ser, allo-Thr, Met, D-Met, Met(O), D-Met(O), Val, D-Val
Tyrosine	Y	D-Tyr, Phe, D-Phe, L-Dopa, His, D-His
Valine	V	D-Val, Leu, D-Leu, Ile, D-Ile, Met, D-Met

Amino acid sequence variants of a protein can be prepared by any of a variety of methods known to those skilled in the art. For example, random mutagenesis of DNA which encodes a protein or a particular domain or region of a protein can be used, e.g., PCR mutagenesis (using, e.g., reduced *Taq* polymerase fidelity to introduce random mutations into a cloned fragment of DNA; Leung et al., Technique 1:11-15 (1989)), or saturation mutagenesis (by, e.g., chemical treatment or irradiation of single-stranded DNA in vitro, and synthesis of a complementary DNA strand; Mayers et al., Science 229:242 (1985)). Random mutagenesis can also be accomplished by, e.g., degenerate oligonucleotide generation (using, e.g., an automatic DNA synthesizer to chemically synthesize degenerate sequences; Narang, Tetrahedron 39:3 (1983); Itakura et al., Recombinant DNA, Proc. 3rd Cleveland Sympos. Macromolecules, ed. A.G. Walton,

Amsterdam: Elsevier, pp. 273-289 (1981)). Non-random or directed mutagenesis can be used to provide specific sequences or mutations in specific regions. These techniques can be used to create variants which include, e.g., deletions, insertions, or substitutions, of residues of the known amino acid sequence of a protein. The sites for mutation can be modified individually or in series, e.g., by (i) substituting first with conserved amino acids and then with more radical choices depending upon results achieved, (ii) deleting the target residue, (iii) inserting residues of the same or a different class adjacent to the located site, or (iv) combinations of the above.

Methods for identifying desirable mutations include, e.g., alanine scanning mutagenesis (Cunningham and Wells, Science 244:1081-1085 (1989)), oligonucleotide-mediated mutagenesis (Adelman et al., DNA 2:183 (1983)); cassette mutagenesis (Wells et al., Gene 34:315 (1985)), combinatorial mutagenesis, and phage display libraries (Ladner et al., WO88/06630).

Other analogs within the invention include, e.g., those with modifications which increase peptide stability. Such analogs may contain, e.g., one or more non-peptide bonds (which replace the peptide bonds) in the peptide sequence. Also included are, e.g.: analogs that include residues other than naturally occurring L-amino acids, e.g., D-amino acids or non-naturally occurring or synthetic amino acids, e.g., β or γ amino acids; and cyclic analogs.

Analogs can be made by methods known to those skilled in the art. For example, analogs can be made by in vitro DNA sequence modifications of the sequence of Fig. 1 (SEQ ID NO:1). For example, in vitro mutagenesis can be used to convert the DNA sequence of Fig. 1 (SEQ ID NO:1) into a sequence which encodes an analog in which one or more amino acid residues has undergone a replacement, e.g., a conservative replacement as described in Table 1.

By fragment is meant some portion of the naturally occurring HNF4 polypeptide. Preferably, the fragment is at least about 20 amino acid residues, more preferably at least about 40 amino acid residues, and most preferably at least about 60 amino acid residues in length. Fragments include, e.g., truncated secreted forms, proteolytic fragments, splicing fragments, other fragments, and chimeric constructs between at least a portion of the relevant gene, e.g., HNF4, and another molecule. Fragments of HNF4 can be generated by methods known to those skilled in the art. The ability of a candidate fragment to exhibit a biological activity of HNF4 can be assessed by methods known to those skilled in the art. Also included are HNF4 fragments containing residues that are not required for biological activity of the fragment or that result from alternative mRNA splicing or alternative protein processing events.

Fragments of a protein can be produced in several ways, e.g., recombinantly, by

proteolytic digestion, or by chemical synthesis. Internal or terminal fragments of a polypeptide can be generated by removing one or more nucleotides from one end (for a terminal fragment) or both ends (for an internal fragment) of a nucleic acid which encodes the polypeptide. Expression of the mutagenized DNA produces polypeptide fragments. Digestion with "end-nibbling" endonucleases can thus generate DNAs which encode an array of fragments. DNAs which encode fragments of a protein can also be generated, e.g., by random shearing, restriction digestion or a combination of the above-discussed methods. For example, fragments of HNF4 can be made by expressing HNF4 DNA which has been manipulated in vitro to encode the desired fragment, e.g., by restriction digestion of the DNA sequence of Fig. 1 (SEQ ID NO:1).

Fragments can also be chemically synthesized using techniques known in the art, e.g., conventional Merrifield solid phase f-Moc or t-Boc chemistry. For example, peptides of the present invention can be arbitrarily divided into fragments of desired length with no overlap of the fragments, or divided into overlapping fragments of a desired length.

HNF4 or a biologically active fragment or analog thereof, or a binding molecule or a biologically active fragment or analog thereof, can, e.g., compete with its cognate molecule for the binding site on the complementary molecule, and thereby reduce or eliminate binding between HNF4 and the cellular binding molecule. HNF4 can be obtained, e.g., from purification or secretion of naturally occurring HNF4 or binding molecule, from recombinant HNF4 or binding molecule, or from synthesized HNF4 or binding molecule.

Therefore, methods for generating analogs and fragments and testing them for activity are known to those skilled in the art.

An agent can also be a nucleic acid used as an antisense molecule. Antisense therapy is meant to include, e.g., administration or in situ generation of oligonucleotides or their derivatives which specifically hybridize, e.g., bind, under cellular conditions, with the cellular mRNA and/or genomic DNA encoding an HNF4 polypeptide, or mutant thereof, so as to inhibit expression of the encoded protein, e.g., by inhibiting transcription and/or translation. The binding may be by conventional base pair complementarity, or, for example, in the case of binding to DNA duplexes, through specific interactions in the major groove of the double helix.

In certain embodiments, the antisense construct binds to a naturally-occurring sequence of an HNF4 gene which, e.g., is involved in expression of the gene. These sequences include, e.g., start codons, stop codons, and RNA primer binding sites.

In other embodiments, the antisense construct binds to a nucleotide sequence which is not

present in the wild type gene. For example, the antisense construct can bind to a region of an HNF4 gene which contains an insertion of an exogenous, non-wild type sequence. Alternatively, the antisense construct can bind to a region of an HNF4 gene which has undergone a deletion, thereby bringing two regions of the gene together which are not normally positioned together and which, together, create a non-wild type sequence. When administered in vivo to a subject, antisense constructs which bind to non-wild type sequences provide the advantage of inhibiting the expression of a mutant HNF4 gene, without inhibiting expression of any wild type HNF4 gene.

An antisense construct of the present invention can be delivered, e.g., as an expression plasmid which, when transcribed in the cell, produces RNA which is complementary to at least a unique portion of the cellular mRNA which encodes an HNF4 polypeptide. An alternative is that the antisense construct is an oligonucleotide probe which is generated ex vivo and which, when introduced into the cell causes inhibition of expression by hybridizing with the mRNA and/or genomic sequences of an HNF4 gene. Such oligonucleotide probes are preferably modified oligonucleotides which are resistant to endogenous nucleases, e.g. exonucleases and/or endonucleases, and are therefore stable in vivo. Exemplary nucleic acid molecules for use as antisense oligonucleotides are phosphoramidate, phosphothioate and methylphosphonate analogs of DNA . (See also U.S. Patents 5,176,996; 5,264,564; and 5,256,775). Additionally, general approaches to constructing oligomers useful in antisense therapy have been reviewed. (See, e.g., Van der Krol et al., Biotechniques 6:958-976, (1988); Stein et al., Cancer Res. 48:2659-2668 (1988)).

By mimetic is meant a molecule which resembles in shape and/or charge distribution HNF4 or a binding molecule. The mimetic can be a peptide or a non-peptide. Mimetics can act as therapeutic agents because they can, e.g., competitively inhibit binding of HNF4 to a binding molecule. By employing, e.g., scanning mutagenesis, e.g., alanine scanning mutagenesis, linker scanning mutagenesis or saturation mutagenesis, to map the amino acid residues of a particular HNF4 polypeptide involved in binding a binding molecule, peptide mimetics, e.g., diazopine or isoquinoline derivatives, can be generated which mimic those residues in binding to a binding molecule, and which therefore can inhibit binding of the HNF4 to a binding molecule and thereby interfere with the function of HNF4. For example, non-hydrolyzable peptide analogs of such residues can be generated using benzodiazepine (see, e.g., Freidinger et al., in Peptides: Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands (1988));

azepine (see, e.g., Huffman et al., in Peptides: Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands (1988)); substituted gama lactam rings (see, e.g., Garvey et al., in Peptides: Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands (1988)); keto-methylene pseudopeptides (see, e.g., Ewenson et al., J. Med. Chem. 29:295 (1986); Ewenson et al., in Peptides: Structure and Function (Proceedings of the 9th American Peptide Symposium) Pierce Chemical Co. Rockland, IL (1985)); β-turn dipeptide cores (see, e.g., Nagai et al., Tetrahedron Lett. 26:647 (1985); Sato et al., J. Chem. Soc. Perkin Trans. 1:1231 (1986)); or β-aminoalcohols (see, e.g., Gordon et al., Biochem. Biophys. Res. Commun. 126:419 (1985); Dann et al., Biochem. Biophys. Res. Commun. 134:71 (1986)).

Antibodies are meant to include antibodies against any moiety that directly or indirectly affects HNF4 metabolism. The antibodies can be directed against, e.g., HNF4 or a binding molecule, or a subunit or fragment thereof. For example, antibodies include anti-HNF4 antibodies; anti-binding molecule antibodies; and Fab_2' fragments of the inhibitory antibody generated through, e.g., enzymatic cleavage. Both polyclonal and monoclonal antibodies can be used in this invention. Preferably, monoclonal antibodies are used. Most preferably, the antibodies have a constant region derived from a human antibody and a variable region derived from an inhibitory mouse monoclonal antibody. Antibodies to HNF4 are described in Pagge Van Strandmann et al., Protein Engineering 8:733-735 (1995).

Agents also include inhibitors of a molecule that are required for synthesis, post-translational modification, or functioning of HNF4 and/or a binding molecule, or activators of a molecule that inhibits the synthesis or functioning of HNF4 and/or the binding molecule. Agents include, e.g., cytokines, growth factors, hormones, signaling components, kinases, phosphatases, homeobox proteins, transcription factors, translation factors and post-translation factors or enzymes. Agents are also meant to include ionizing radiation, non-ionizing radiation, ultrasound and toxic agents which can, e.g., at least partially inactivate or destroy HNF4 and/or the binding molecule.

An agent is also meant to include agents which are not entirely HNF4 specific. For example, an agent may alter other insulin metabolism related genes or proteins. Such overlapping specificity may provide additional therapeutic advantage.

In certain embodiments, the method employs two phases for evaluating an agent for use in treating diabetes, an initial in vitro phase and then an in vivo phase. The agent is administered to the test cell or cell-free system in vitro, and if a change in an aspect of HNF4 metabolism

occurs, then the agent is further administered to a test animal in a therapeutically effective amount and evaluated in vivo for an effect of the agent on an aspect of HNF4 metabolism.

The invention also includes the agent so identified as being useful in treating diabetes.

The invention also includes a method for evaluating an agent for the ability to alter the binding of HNF4 polypeptide to a binding molecule. An agent is provided. An HNF4 polypeptide is provided. A binding molecule is provided. The agent, HNF4 polypeptide and binding molecule are combined. The formation of a complex comprising the HNF4 polypeptide and binding molecule is detected. An alteration in the formation of the complex in the presence of the agent as compared to in the absence of the agent is indicative of the agent altering the binding of the HNF4 polypeptide to the binding molecule.

Altering the binding includes, e.g., inhibiting or promoting the binding. The efficacy of the agent can be assessed, e.g., by generating dose response curves from data obtained using various concentrations of the agent. Methods for determining formation of a complex are standard and are known to those skilled in the art.

The invention also includes the agent so identified as being able to alter the binding of HNF4 polypeptide to a binding molecule.

The invention also includes a method for evaluating an agent for the ability to bind to HNF4 polypeptide. An agent is provided. An HNF4 polypeptide is provided. The agent is contacted with the HNF4 polypeptide. The ability of the agent to bind to the HNF4 polypeptide is evaluated. Binding can be determined, e.g., by measuring formation of a complex by standard methods known to those skilled in the art.

The invention also includes the agent so identified as being able to bind to HNF4 polypeptide.

The invention also includes a method for evaluating an agent for the ability to bind to a nucleic acid encoding an HNF4 regulatory sequence. An agent is provided. A nucleic acid encoding an HNF4 regulatory sequence is provided. The agent is contacted with the nucleic acid. The ability of the agent to bind to the nucleic acid is evaluated. Binding can be determined, e.g., by measuring formation of a complex by standard methods known to those skilled in the art.

The invention also includes the agent so identified as being able to bind to a nucleic acid encoding an HNF4 regulatory sequence.

The invention also includes a method for treating diabetes in an animal. An animal in

need of treatment for diabetes is provided. An agent capable of altering an aspect of HNF4 structure or metabolism is provided. The agent is administered to the animal in a therapeutically effective amount such that treatment of the diabetes occurs.

Treating is meant to include, e.g., preventing, treating, reducing the symptoms of, or curing the diabetes. Administration of the agent can be accomplished by any method which allows the agent to reach the target cells. These methods include, e.g., injection, deposition, implantation, suppositories, oral ingestion, inhalation, topical administration, or any other method of administration where access to the target cells by the agent is obtained. Injections can be, e.g., intravenous, intradermal, subcutaneous, intramuscular or intraperitoneal. Implantation includes inserting implantable drug delivery systems, e.g., microspheres, hydrogels, polymeric reservoirs, cholesterol matrices, polymeric systems, e.g., matrix erosion and/or diffusion systems and non-polymeric systems, e.g., compressed, fused or partially fused pellets. Suppositories include glycerin suppositories. Oral ingestion doses can be enterically coated. Inhalation includes administering the agent with an aerosol in an inhalator, either alone or attached to a carrier that can be absorbed.

Administration of the agent can be alone or in combination with other therapeutic agents. In certain embodiments, the agent can be combined with a suitable carrier, incorporated into a liposome, or incorporated into a polymer release system.

In certain embodiments of the invention, the administration can be designed so as to result in sequential exposures to the agent over some time period, e.g., hours, days, weeks, months or years. This can be accomplished by repeated administrations of the agent by one of the methods described above, or alternatively, by a controlled release delivery system in which the agent is delivered to the animal over a prolonged period without repeated administrations. By a controlled release delivery system is meant that total release of the agent does not occur immediately upon administration, but rather is delayed for some time period. Release can occur in bursts or it can occur gradually and continuously. Administration of such a system can be, e.g., by long acting oral dosage forms, bolus injections, transdermal patches or sub-cutaneous implants.

Examples of systems in which release occurs in bursts include, e.g., systems in which the agent is entrapped in liposomes which are encapsulated in a polymer matrix, the liposomes being sensitive to a specific stimuli, e.g., temperature, pH, light or a degrading enzyme, and systems in which the agent is encapsulated by an ionically-coated microcapsule with a microcapsule

core-degrading enzyme. Examples of systems in which release of the agent is gradual and continuous include, e.g., erosional systems in which the agent is contained in a form within a matrix, and diffusional systems in which the agent permeates at a controlled rate, e.g., through a polymer. Such sustained release systems can be, e.g., in the form of pellets or capsules.

The agent can be suspended in a liquid, e.g., in dissolved form or colloidal form. The liquid can be a solvent, partial solvent or non-solvent. In many cases water or an organic liquid can be used.

The agent can be administered prior to or subsequent to the appearance of diabetic symptoms. In certain embodiments, the agent is administered to patients with familial histories of diabetes, or who have phenotypes that may indicate a predisposition to diabetes, or who have been diagnosed as having a genotype which predisposes the patient to diabetes.

The agent is administered to the animal in a therapeutically effective amount. By therapeutically effective amount is meant that amount which is capable of at least partially preventing or reversing diabetes. A therapeutically effective amount can be determined on an individual basis and will be based, at least in part, on consideration of the species of animal, the animal's size, the animal's age, the agent used, the type of delivery system used, the time of administration relative to the onset of diabetic symptoms, and whether a single, multiple, or controlled release dose regimen is employed. A therapeutically effective amount can be determined by one of ordinary skill in the art employing such factors and using no more than routine experimentation.

Preferably, the concentration of the agent is at a dose of about 0.1 to about 1000 mg/kg body weight, more preferably at about 0.1 to about 500 mg/kg, more preferably yet at about 0.1 to about 100 mg/kg, and most preferably at about 0.1 to about 5 mg/kg. The specific concentration partially depends upon the particular agent used, as some are more effective than others. The dosage concentration of the agent that is actually administered is dependent at least in part upon the final concentration that is desired at the site of action, the method of administration, the efficacy of the particular agent, the longevity of the particular agent, and the timing of administration relative to the onset of the diabetic symptoms. Preferably, the dosage form is such that it does not substantially deleteriously affect the animal. The dosage can be determined by one of ordinary skill in the art employing such factors and using no more than routine experimentation.

In certain embodiments, various gene constructs can be used as part of a gene therapy

protocol to deliver nucleic acids encoding, e.g., either an agonistic or antagonistic form of an HNF4 polypeptide. Expression vectors can be used for in vivo transfection and expression of an HNF4 polypeptide in particular cell types so as to reconstitute the function of, or alternatively, abrogate the function of HNF4 polypeptide in a cell in which non-wild type HNF4 is expressed. Expression constructs of the HNF4 polypeptide, and mutants thereof, may be administered in any biologically effective carrier, e.g. any formulation or composition capable of effectively delivering the HNF4 gene to cells in vivo. Approaches include, e.g., insertion of the subject gene in viral vectors including, e.g., recombinant retroviruses, adenovirus, adeno-associated virus, and herpes simplex virus-1, or recombinant bacterial or eukaryotic plasmids. Viral vectors transfect cells directly; plasmid DNA can be delivered with the help of, for example, cationic liposomes (lipofectin) or derivatized (e.g. antibody conjugated), polylysine conjugates, gramacidin S, artificial viral envelopes or other such intracellular carriers, as well as direct injection of the gene construct or $CaPO_4$ precipitation carried out in vivo. The above-described methods are known to those skilled in the art and can be performed without undue experimentation. Since transduction of appropriate target cells represents the critical first step in gene therapy, choice of the particular gene delivery system will depend on such factors as the phenotype of the intended target and the route of administration, e.g., locally or systemically. Administration can be directed to one or more cell types, and to one or more cells within a cell type, so as to be therapeutically effective, by methods that are known to those skilled in the art. In a preferred embodiment, the agent is administered to beta cells of the animal. For example, a genetically engineered HNF4 gene is administered to beta cells. In certain embodiments, administration is done in a prenatal animal or embryonic cell. It will be recognized that the particular gene construct provided for in in vivo transduction of HNF4 expression are also useful for in vitro transduction of cells, such as for use in the diagnostic assays described above.

The invention also includes a method for treating an animal at risk for diabetes. An animal at risk for diabetes is provided. An agent capable of altering an aspect of HNF4 structure or metabolism is provided. The agent is administered to the animal in a therapeutically effective amount such that treatment of the animal occurs. Being at risk for diabetes can result from, e.g., a familial history of diabetes, phenotypic symptoms which predispose to diabetes, or a genotype which predisposes to diabetes.

The invention also includes a method for treating a cell having an abnormality in structure or metabolism of HNF4. A cell having an abnormality in structure or metabolism of

HNF4 is provided. An agent capable of altering an aspect of HNF4 structure or metabolism is provided. The agent is administered to the cell in a therapeutically effective amount such that treatment of the cell occurs.

In certain embodiments, the cell is obtained from a cell or tissue culture or an embryo fibroblast. The cell can be, e.g., part of an animal, e.g., a natural animal or a non-human transgenic animal.

The invention also includes a pharmaceutical composition for treating diabetes in an animal comprising a therapeutically effective amount of an agent, the agent being capable of altering an aspect of HNF4 metabolism or structure in the animal so as to result in treatment of the diabetes, and a pharmaceutically acceptable carrier.

The invention also includes a vaccine composition for treating diabetes in an animal comprising a therapeutically effective amount of an agent, the agent being capable of altering an aspect of HNF4 metabolism or structure in the animal so as to result in treatment of the diabetes, and a pharmaceutically acceptable carrier.

The invention also includes a method of making a fragment or analog of HNF4 polypeptide, the fragment or analog having a transcription activating activity of a natually-occurring HNF4 polypeptide for a gene involved in insulin metabolism. An HNF4 polypeptide is provided. The sequence of the HNF4 polypeptide is altered. The altered HNF4 polypeptide is tested for the transcription activating activity. The fragments or analogs can be generated and tested for the presence of HNF4 transcription activating activity by methods known to those skilled in the art, e.g., as described herein.

The invention also includes a method of making an HNF4 polypeptide or fragment or analog thereof having a non-wild type activity as a transcription activator for a gene involved in insulin metabolism. The sequence of an HNF4 polypeptide is altered. The altered HNF4 polypeptide is tested for the non-wild type activity.

By a non-wild type activity as a transcription activator is meant an activity which differs from wild type. The activity can be greater or it can be less, including substantially no detectable activity. In certain embodiments, the altered HNF4 polypeptide possesses substantially about 0%, or at least about 10%, or at least about 40%, or at least about 90% of the activity of wild type HNF4 in an in vivo or an in vitro assay.

The following non-limiting examples further illustrate the present invention.

EXAMPLES

Example 1: Linkage Analysis of Common Form NIDDM with Markers in the MODY 1 Region

This example illustrates linkage studies on families with common form NIDDM which show that a gene within the MODY 1 region on the long arm of human chromosome 20 (20q) contributes to the development of the common form of NIDDM in a significant number of diabetic families. This region includes the HNF4 gene.

Linkage studies were conducted in 26 Caucasian families in which NIDDM had been diagnosed between 30 and 60 years of age and in which it appeared to segregate as an autosomal dominant disorder. A total of 449 individuals, including 129 NIDDM patients, were genotyped for 8 highly polymorphic microsatellite markers spanning a 31 cM region on chromosome 20q12-13.1. Using affected sib-pair and robust sib-pair analyses, evidence was obtained for linkage between NIDDM and markers D20S119, D20S178, and D20S197. These markers span a chromosomal region that has been shown to segregate with MODY in the R.W. family. Based upon the posterior probability of linkage, 27% of the examined families exhibited linkage to the MODY 1 region. In these families, evidence in favor of linkage was quite strong for the markers in the region supported by sib-pair analysis, with maximum lod score occurring at θ=0.00 for D20S119 (lod score = 2.86), at θ=0.00 for D20S178 (lod score = 2.86), and θ=0.01 for D20S197 (lod score = 1.88). Multipoint non-parametric linkage analysis showed the strongest linkage of NIDDM with D20S178/D20S197 (Z score of 5.3, p=0.0001).

The 26 NIDDM pedigrees that were examined were ascertained through the following three characteristics: a proband with NIDDM diagnosed between age 30 and 59 years, NIDDM present in at least one sibling of the proband, and one and only one parent having diabetes according to the information available to the proband. The 449 individuals examined ranged from 15-95 years of age. On average, 5 NIDDM individuals in one or two generations were examined in each family (range from 2 to 9). In addition, an average of 12 non-diabetic individuals (usually in 2 or 3 generations) were examined in each family (range 3 to 29). The mean age of the non-diabetic individuals at the time of examination was similar to the age at onset of NIDDM in their relatives. The non-diabetic individuals were selected because they were siblings of NIDDM individuals or were children or grandchildren of deceased NIDDM individuals. On examining the families more carefully, it was found that four of them were bilineal, whereas in the others, diabetes occurred on only one side of family in at least two

generations.

Comparisons of the examined individuals according to diabetes status were made. Diabetes had been diagnosed in 129 subjects, 25 individuals had impaired glucose tolerance and 295 were non-diabetics. On average, the patients with NIDDM had diabetes diagnosed 12 years prior to this examination. At the time of examination, 45% of them were treated with insulin, 31% were treated with oral agents, and the remaining 24% were treated with diet or had been untreated because diabetes had been undetected until this examination (n=21). The subjects with diabetes had been very obese in the past (their lifetime maximum-body-wejght averaged more than 150 percent of ideal) and were also more obese at the time of examination than their non-diabetic relatives.

Based on the current Cooperative Human Linkage Center (CHLC) database (University of Iowa, Iowa City, IA 52242), 8 markers were selected for this study which span a 31 cM region including the MODY 1 region (see Fig. 1). The MODY 1 locus had been mapped to a 13 cM (sex-averaged) region located telomeric to the ADA locus on chromosome 20q12-13.1 (see Fig. 1). Genotyping for all markers was performed by PCR-based protocols (Weber et al., Am. Hum. Genet. 44:388-396 (1989)). Generally, the forward PCR primer was end-labeled with (32γ)-P-ATP by T4 polynucleotide kinase (Promega, Madison, WI) according to the manufacturer's protocol. PCR was performed with a 10 μl volume in a 96-well PCR plate (MJ Research, Watertown, MA). Each PCR reaction contained 50 ng of genomic DNA, 10 pmol of each primer, 25 μM dNTPs, 0.3-0.5 mM $MgCl_2$ and 0.2 units of *Taq* polymerase (Roche Molecular Systems, Inc., Branchburg, New Jersey) with 1μl of 10xPCR buffer. All PCR reactions were performed following the "hot-start" procedure by adding *Taq* polymerase and labeled primer mixture (in a total volume of 2μl) after a denaturation step of 94°C for 5 minutes. Samples were subjected to 30 cycles of 40 seconds at 94°C for denaturing, 1 minute at optimum annealing temperature, and 1 minute at 72°C for elongation, using a 96-well Thermocycler (MJ Research, Watertown, MA). PCR reactions ended with 5 minutes of final incubation at 72°C. After PCR amplification, 2 volumes of sequencing stop solution was added to the PCR reaction, the mixture was heated at 95°C for 5 minutes, 2 μL of the denatured PCR product was loaded onto 5% polyacrylamide DNA sequencing gels and resolved electrophoretically in 0.6 x TBE at 100-W constant current. Consistency between different gels and between different lanes of the same gel was maintained by using either standard PCR products of known genotype or size standards from Research Genetics (Huntsville, AL). Dried gels were exposed to X-ray films for 12 hours

without using the intensifying screen. Autoradiograms were scored by two independent observers. PCR primers were purchased from Research Genetics (Huntsville, AL). Primer sequences were available from Genome Data Base (Johns Hopkins University). Allele frequencies of markers were estimated from 80 unrelated non-diabetic individuals (the majority being spouses of NIDDM patients).

After genotyping 449 individuals for 8 highly polymorphic microsatellites, the genotypes of an additional 69 non-participating individuals (the majority of them dead) could be inferred, including 39 diabetic subjects. Using all of the observed and inferred genotypes, a total of 165 to 173 (depending on the genetic markers) affected sib-pairs were constructed. The proportion of alleles shared IBD by affected sib-pairs for each marker was determined, and is shown in Table 2.

Table 2

RESULT OF SIB-PAIR ANALYSES

Marker	Heterozygosity of marker	Affected sib-pair Analysis*		Robust sib-pair Analysis†	
		Mean sharing of IBD alleles	p	Slope	p
D20S107	0.80	0.52	0.184	-0.148	0.011
ADA8PR	0.80	0.52	0.168	-0.153	0.011
D20S119	0.82	0.57	0.0008	-0.206	0.002
D20S178	0.83	0.54	0.031	-0.179	0.004
D20S197	0.76	0.53	0.077	-0.197	0.004
D20S176	0.63	0.54	0.031	-0.100	0.103
D20S196	0.81	0.53	0.138	-0.075	0.120
D20S100	0.76	0.52	0.241	-0.088	0.108

* Number of affected sib-pairs varied from 165 to 173.
† Number of total sib-pairs varied from 533 to 549.

The mean proportion of alleles shared IBD was higher than 0.50 for every marker studied; however, the excess in IBD sharing was significantly greater than 0.5 only for D20S119, D20S178, and D20S176, all telomeric to the ADA locus. The result of robust sib-pair analysis is also shown in Table 2. As in the affected sib-pair analysis, markers D20S119 and D20S178 exhibited strong evidence of linkage with NIDDM, as did D20S197, but not D20S176.

Additional markers (D20S107 and ADA8PR) also showed linkage with NIDDM in this analysis. The results of robust sib-pair analysis did not change after adjustment for the age at onset of NIDDM.

Linkage analysis was performed for all families studied, and the lod scores generated for these families are shown in Table 3.

TABLE 3

LINKAGE ANALYSIS OF NIDDM AND CHROMOSOME 20 MARKERS

	All families		"Linked" families*	
Marker Locus	θ	lod	θ	lod
D20S107 (6.1)†	0.30	0.09	0.27	0.28
ADA8PR (1.9)	0.4	0.03	0.24	0.36
D20S119 (5.0)	0.5	0.00	0.00	2.86
D20S178 (0.0)	0.26	0.46	0.00	2.86
S20S197 (5.0)	0.24	0.74	0.01	1.88
S20S176 (4.0)	0.5	0.00	0.08	0.44
D20S196 (9.0)	0.20	0.40	0.09	1.34
D20S100	0.5	0.00	0.50	0.00

*7 families with 125 participants, 32 having NIDDM. NIDDM patients in these and the other families did not differ with regard to age at onset or treatment of NIDDM, or % IBW.

† Sex-averaged genetic distance between markers in cM.

While the overall results are not suggestive of linkage, analysis of the lod scores by family using HOMOG (Ott, J., Am. Hum. Genet. 47:311-320 (1983)), rejected the homogeneity hypothesis ($p<0.05$) for D20S119, D20S178 and D20S197. Based upon the posterior probability of linkage, seven (27%) of these families exhibited linkage to the MODY 1 region. In these families, the evidence in favor of linkage was quite strong for the loci in the regions supported by sib pair analysis, with maximum lod scores obtained at $\theta = 0.00$ for D20S119 (lod score = 2.86), at $\theta = 0.00$ for D20S178 (lod score = 2.86), and at $\theta = 0.01$ for D20S197 (lod score = 1.88).

To examine the consistency of the results of lod score analysis with the results of the sib-pair approach, the 26 families were partitioned into two groups (7 "linked" and 19 "unlinked"

families for a re-examination by sib-pair analysis. In the 7 "linked" families, the mean proportion of alleles at D20S119 shared IBD by affected sib-pairs was 0.70 ($p<0.000\ 1$). In the 19 "unlinked" families, the mean sharing was 0.50, the expected value under the null hypothesis of no linkage. Similar results were obtained for robust sib-pair analysis.

Multipoint linkage analysis was employed to incorporate information from all chromosome 20 markers that were evaluated in the MODY1 region in order to infer more effectively the probability distribution of IBD status at each point in the interval of interest. (See Fig. 2). Fig. 2 shows the results of multipoint nonparametric linkage analysis of NIDDM with markers in the MODY 1 region of chromosome 20q. Results were obtained with GENEHUNTER software (Kruglyak et al., Am. J. Hum. Genet. 58:1347-1363 (1996). On the x-axis, the genetic distance between markers is presented in cM. The MODY 1 locus in the W-R family lies between the ADA8PR locus and the D20S176 marker. (See Bowden et al., Am. J. Hum. Genet. 50:607-618 (1992); Bowden et al., Diabetes 41:88-92 (1992); Rothschild et al., Genomics 13:560-564 (1992); Rothschild et al., Am. J. Hum. Genet. 52:110-123 (1993)). Using multipoint analysis in all families, the peak lod scores occurred for markers D20S119 (Z score 1.8) and D20S197/D20S178 (Z score 1.8). From previous parametric linkage results, there was clear evidence for linkage heterogeneity. Thus, multipoint analyses were repeated in those families that suggested linkage to the MODY1 region and in those that suggested absence of linkage to this region. In the linked families, the multipoint analysis clearly identified the D20S197/D20S178 markers as contributing to NIDDM susceptibility, with a peak Z score of 5.3 ($p=0.0001$); there was no evidence for linkage in the unlinked families.

Evidence for association between NIDDM and D20S119 (the most informative and strongly linked marker) was determined by use of the transmission disequilibrium test (TDT). Using the TDT, there were 195 transmission from a heterozygous parent. There was no significant evidence for deviation of transmitted alleles from that expected, suggesting that while there was evidence for linkage (based upon both sib-pair and linkage analyses) there was no support for association with alleles at D20S119.

The protocol for this study was approved by the Human Subjects Committee of the Joslin Diabetes Center.

Those skilled in the art will be able to ascertain using no more than routine experimentation, many equivalents of the specific embodiments of the invention described herein. These and all other equivalents are intended to be encompassed by the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: Krolewski, Andrzej S.

(ii) TITLE OF THE INVENTION: METHODS FOR DIAGNOSING AND TREATING DIABETES

(iii) NUMBER OF SEQUENCES: 4

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Banner & Witcoff, Ltd.
(B) STREET: One Financial Center
(C) CITY: Boston
(D) STATE: MA
(E) COUNTRY: USA
(F) ZIP: 02111

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM PC Compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release 1.0, Version 1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: Not available
(B) FILING DATE: November 25, 1997
(C) CLASSIFICATION: Not available

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Greer, Helen
(B) REGISTRATION NUMBER: 36,816
(C) REFERENCE/DOCKET NUMBER: 3984/59693

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 617-345-9100
(B) TELEFAX: 617-345-9111

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2288 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GGGGCGGAGG GCGGGGGCTT CGGGGTGGGC GCCCAGGGTA GGGCAGGTGG CCGCCGCGTG 60
GAGGCAGGGA GAATGCACTC TCCAAAACCC TCCGTCGACA TGGACATGGC CGACTACAGT 120
GCTGCACTGG ACCCAGCCTA CACCACCCTG GAATTTGAGA ATGTGCAGGT GTTGACGATG 180
GGCAATGACA CGTCCCCATC AGAAGGCACC AACCTCAACG CGCCCAACAG CCTGGGTGTC 240

AGCGCCCTGT GTGCCATCTG CGGGGACCGG GCCACGGGCA AACACTACGG TGCCTCGAGC 300
TGTGACGGCT GCAAGGGCTT CTTCCGGAGG AGCGTGCGGA AGAACCACAT GTACTCCTGC 360
AGATTTAGCC GGCAGTGCGT GGTGGACAAG GACAAGAGGA ACCAGTGCCG CTACTGCAGG 420
CTCAAGAAAT GCTTCCGGGC TGGCATGAAG AAGGAAGCCG TCCAGAATGA GCGGGACCGG 480
ATCAGCACTC GAAGGTCAAG CTATGAGGAC AGCAGCCTGC CCTCCATCAA TGCGCTCCTG 540
CAGGCGGAGG TCCTGTCCCG ACAGATCACC TCCCCCGTCT CCGGGATCAA CGGCGACATT 600
CGGGCGAAGA AGATTGCCAG CATCGCAGAT GTGTGTGAGT CCATGAAGGA GCAGCTGCTG 660
GTTCTCGTTG AGTGGGCCAA GTACATCCCA GCTTTCTGCG AGCTCCCCCT GGACGACCAG 720
GTGGCCCTGC TCAGAGCCCA TGCTGGCGAG CACCTGCTGC TCGGAGCCAC CAAGAGATCC 780
ATGGTGTTCA AGGACGTGCT GCTCCTAGGC AATGACTACA TTGTCCCTCG GCACTGCCCG 840
GAGCTGGCGG AGATGAGCCG GGTGTCCATA CGCATCCTTG ACGAGCTGGT GCTGCCCTTC 900
CAGGAGCTGC AGATCGATGA CAATGAGTAT GCCTACCTCA AAGCCATCAT CTTCTTTGAC 960
CCAGATGCCA AGGGGCTGAG CGATCCAGGG AAGATCAAGC GGCTGCGTTC CCAGGTGCAG 1020
GTGAGCTTGG AGGACTACAT CAACGACCGC CAGTATGACT CGCGTGGCCG CTTTGGAGAG 1080
CTGCTGCTGC TGCTGCCCAC CTTGCAGAGC ATCACCTGGC AGATGATCGA GCAGATCCAG 1140
TTCATCAAGC TCTTCGGCAT GGCCAAGATT GACAACCTGT TGCAGGAGAT GCTGCTGGGA 1200
GGGTCCCCCA GCGATGCACC CCATGCCCAC CACCCCCTGC ACCCTCACCT GATGCAGGAA 1260
CATATGGGAA CCAACGTCAT CGTTGCCAAC ACAATGCCCA CTCACCTCAG CAACGGACAG 1320
ATGTCCACCC CTGAGACCCC ACAGCCCTCA CCGCCAGGTG GCTCAGGGTC TGAGTCCTAT 1380
AAGCTCCTGC CGGGAGCCGT CGCCACAATC GTCAAGCCCC TCTCTGCCAT CCCCCAGCCG 1440
ACCATCACCA AGCAGGAAGT TATCTAGCAA GCCGCTGGGG CTTGGGGGCT CCACTGGCTC 1500
CCCCCAGTCC CCCTAAGAGA GCACCTGGTG ATACCGTGGT CACGGCAAAG GAAGACGTGA 1560
TGCCAGGACC AGTCCCAGAG CAGGAATGGG AAGGATGAAG GGCCCGAGAA CATGGCCTAA 1620
GGGCACATCC CACTGCCACC CTTGACGCCC CTGCTCTGGA TAACAAGACT TTGACTTGGG 1680
GAGACCTCTA CTGCCTTGGA CAACTTTTCT CATGTTGAAG CCACTGCCTT CACCTTCACC 1740
TTCATCCATG TCCAACCCCC GACTTCATCC CAAAGGACAG CCGCCTGGAG ATGACTTGAG 1800
GCCTTACTTA AACCCAGCTC CCTGTGCCCT AGCCTGGTGC TTCTCCTCTC CTAGCCCCTG 1860
TCATGGTGTC CCAGACAGAG CCCTGTGAGG CTGGGTCCAA TTGTGGCACT TGGGGCACCT 1920
TGCTCCTCCT TCTGCTGCTG CCCCCACCTC TGCTGTCACC TTGCTCAGCC ATCCCGTCTT 1980
CTCCAACACC ACCTCTCCAG AGGCCAAGGA GGCCTTGGAA ACGATTCCCC CAGTCATTCT 2040
GGGAATACTG TTGTAAGACC TGACTGGGCC AGGCACCAGG CAGGGTCTAG AAGGCTGTGG 2100
TGAGGGAAGA CGCCTTTCTC CTCCAACCCA ACCTCATCCT CCTTCTTCAG GGACTTGGGT 2160
GGTACTTGGG TGAGGATCCC TAAGGCCTTC AACCCGAGAA AACAAACCCA GGTTGGCGAC 2220
TGCAACAGGA ACTTGGAGTG GAGAGGAAAA GCATCAGAAA GAGGCAGACC ATCCACCAGG 2280
CCTTTGAG 2288

(3) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1635 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GGCTTCGGGG TGGGCGCCCA GGGTAGGGCA GGTGGCCGCC GCGTGGAGGC AGGGAGAATG 60
CACTCTCCAA AACCCTCCGT CGACATGGAC ATGGCCGACT ACAGTGCTGC ACTGGACCCA 120
GCCTACACCA CCCTGGAATT TGAGAATGTG CAGGTGTTGA CGATGGGCAA TGACACGTCC 180
CCATCAGAAG GCACCAACCT CAACGCGCCC AACAGCCTGG GTGTCAGCGC CCTGTGTGCC 240
ATCTGCGGGG ACCGGGCCAC GGGCAAACAC TACGGTGCCT CGAGCTGTGA CGGCTGCAAG 300
GGCTTCTTCC GGAGGAGCGT GCGGAAGAAC CACATGTACT CCTGCAGATT TAGCCGGCAG 360

TGCGTGGTGG ACAAGGACAA GAGGAACCAG TGCCGCTACT GCAGGCTCAA GAAATGCTTC 420
CGGGCTGGCA TGAAGAAGGA AGCCGTCCAG AATGAGCGGG ACCGGATCAG CACTCGAAGG 480
TCAAGCTATG AGGACAGCAG CCTGCCCTCC ATCAATGCGC TCCTGCAGGC GGAGGTCCTG 540
TCCCGACAGA TCACCTCCCC CGTCTCCGGG ATCAACGGCG ACATTCGGGC GAAGAAGATT 600
GCCAGCATCG CAGATGTGTG TGAGTCCATG AAGGAGCAGC TGCTGGTTCT CGTTGAGTGG 660
GCCAAGTACA TCCCAGCTTT CTGCGAGCTC CCCCTGGACG ACCAGGTGGC CCTGCTCAGA 720
GCCCATGCTG GCGAGCACCT GCTGCTCGGA GCCACCAAGA GATCCATGGT GTTCAAGGAC 780
GTGCTGCTCC TAGGCAATGA CTACATTGTC CCTCGGCACT GCCCGGAGCT GGCGGAGATG 840
AGCCGGGTGT CCATACGCAT CCTTGACGAG CTGGTGCTGC CCTTCCAGGA GCTGCAGATC 900
GATGACAATG AGTATGCCTA CCTCAAAGCC ATCATCTTCT TTGACCCAGA TGCCAAGGGG 960
CTGAGCGATC CAGGGAAGAT CAAGCGGCTG CGTTCCCAGG TGCAGGTGAG CTTGGAGGAC 1020
TACATCAACG ACCGCCAGTA TGACTCGCGT GGCCGCTTTG GAGAGCTGCT GCTGCTGCTG 1080
CCCACCTTGC AGAGCATCAC CTGGCAGATG ATCGAGCAGA TCCAGTTCAT CAAGCTCTTC 1140
GGCATGGCCA AGATTGACAA CCTGTTGCAG GAGATGCTGC TGGGAGGGTC CCCCAGCGAT 1200
GCACCCCATG CCCACCACCC CCTGCACCCT CACCTGATGC AGGAACATAT GGGAACCAAC 1260
GTCATCGTTG CCAACACAAT GCCCACTCAC CTCAGCAACG GACAGATGTG TGAGTGGCCC 1320
CGACCCAGGG GACAGGCAGC CACCCCTGAG ACCCCACAGC CCTCACCGCC AGGTGGCTCA 1380
GGGTCTGAGT CCTATAAGCT CCTGCCGGGA GCCGTCGCCA CAATCGTCAA GCCCCTCTCT 1440
GCCATCCCCC AGCCGACCAT CACCAAGCAG GAAGTTATCT AGCAAGCCGC TGGGGCTTGG 1500
GGGCTCCACT GGCTCCCCCC AGTCCCCCTA AGAGAGCACC TGGTGATACC GTGGTCACGG 1560
CAAAGGAAGA CGTGATGCCA GGACCAGTCC CAGAGCAGGA ATGGGAAGGA TGAAGGGCCC 1620
GAGAACATGG CCTAA 1635

(4) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2289 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GCAGAGAGGG CACTGGGAGG AGGCAGTGGG AGGGCGGAGG GCGGGGGCCT TCGGGGTGGG 60
CGCCCAGGGT AGGGCAGGTG GCCGCGGCGT GGAGGCAGGG AGAATGCGAC TCTCCAAAAC 120
CCTCGTCGAC ATGGACATGG CCGACTACAG TGCTGCACTG GACCCAGCCT ACACCACCCT 180
GGAATTTGAG AATGTGCAGG TGTTGACGAT GGGCAATGAT TTGTTGCCGC TGCGTCTCGC 240
CAGATTGAGG CATCCCCTCC GACATCACTG GAGCATATCT GGAGGGGTGG ACAGTTCTCC 300
ACAGGGAGAC ACGTCCCCAT CAGAAGGCAC CAACCTCAAC GCGCCCAACA GCCTGGGTGT 360
CAGCGCCCTG TGTGCCATCT GCGGGGACCG GGCCACGGGC AAACACTACG GTGCCTCGAG 420
CTGTGACGGC TGCAAGGGCT TCTTCCGGAG GAGCGTGCGG AAGAACCACA TGTACTCCTG 480
CAGATTTAGC CGGCAGTGCG TGGTGGACAA AGACAAGAGG AACCAGTGCC GCTACTGCAG 540
GCTCAAGAAA TGCTTCCGGG CTGGCATGAA GAAGGAAGCC GTCCAGAATG AGCGGGACCG 600
GATCAGCACT CGAAGGTCAA GCTATGAGGA CAGCAGCCTG CCCTCCATCA ATGCGCTCCT 660
GCAGGCGGAG GTCCTGTCCC GACAGATCAC CTCCCCCGTC TCCGGGATCA ACGGCGACAT 720
TCGGGCGAAG AAGATTGCCA GCATCGCAGA TGTGTGTGAG TCCATGAAGG AGCAGCTGCT 780
GGTTCTCGTT GAGTGGGCCA AGTACATCCC AGCTTTCTGC GAGCTCCCCC TGGACGACCA 840
GGTGGCCCTG CTCAGAGCCC ATGCTGGCGA GCACCTGCTG CTCGGAGCCA CCAAGAGATC 900
CATGGTGTTC AAGGACGTGC TGCTCCTAGG CAATGACTAC ATTGTCCCTC GGCACTGCCC 960
GGAGCTGGCG GAGATGAGCC GGGTGTCCAT ACGCATCCTT GACGAGCTGG TGCTGCCCTT 1020
CCAGGAGCTG CAGATCGATG ACAATGAGTA TGCCTACCTC AAAGCCATCA TCTTCTTTGA 1080
CCCAGATGCC AAGGGGCTGA GCGATCCAGG GAAGATCAAG CGGCTGCGTT CCCAGGTGCA 1140
GGTGAGCTTG GAGGACTACA TCAACGACCG CCAGTATGAC TCGCGTGGCC GCTTTGGAGA 1200

GCTGCTGCTG CTGCTGCCCA CCTTGCAGAG CATCACGTGG CAGATGATCG AGCAGATCCA 1260
GTTCATCAAG CTCTTCGGCA TGGCCAAGAT TGACAACCTG TTGCAGGAGA TGCTGCTGGG 1320
AGGGTCCCCC AGCGATGCAC CCCATGCCCA CCACCCCCTG CACCCTCACC TGATGCAGGA 1380
ACATATGGGA ACCAACGTCA TCGTTGCCAA CACAATGCCC ACTCACCTCA GCAACGGACA 1440
GATGTGTGAG TGGCCCCGAC CCAGGGGACA GGCAGCCACC CCTGAGACCC CACAGCCCTC 1500
ACCGCCAGGT GGCTCAGGGT CTGAGCCCTA TAAGCTCCTG CCGGGAGCCG TCGCCACAAT 1560
CGTCAAGCCC CTCTCTGCCA TCCCCCAGCC GACCATCACC AAGCAGGAAG TTATCTAGCA 1620
AGCCGCTGGG GCTTGGGGGC TCCACTGGCT CCCCCCAGCC CCCTAAGAGA GCACCTGGTG 1680
ATCACGTGGT CACGGCAAAG GAAGACGTGA TGCCAGGACC AGTCCCAGAG CAGGAATGGG 1740
AAGGATGAAG GGCCCGAGAA CATGGCCTAA GGCACATCCC ACTGCACCCT GACGCCCTGC 1800
TCTGATAACA AGACTTTGAC TTGGGGAGAC CCTCTACTGC CTTGGACAAC TTTCTCATGT 1860
TGAAGCCACT GCCTTCACCT TCACCTTCAT CCATGTCCAA CCCCCGACTT CATCCCAAAG 1920
GACAGCCGCC TGGAGATGAC TTGAGCCTTA CTTAAACCCA GCTCCCTTCT TCCCTAGCCT 1980
GGTGCTTCTC CTCTCCTAGC CCCGGTCATG GTGTCCAGAC AGAGCCCTGT GAGGCTGGGT 2040
CCAATTGTGG CACTTGGGGC ACCTTGCTCC TCCTTCTGCT GCTGCCCCCA CCTCTGCTGC 2100
CTCCCTCTGC TGTCACCTTG CTCAGCCATC CCGTCTTCTC CAACACCACC TCTACAGAGG 2160
CCAAGGAGGC CTTGGAAACG ATTCCCCCAG TCATTCTGGG AACATGTTGT AAGCACTGAC 2220
TGGGACCAGG CACCAGGCAG GGTCTAGAAG GCTGTGGTGA GGGAAGACGC CTTTCTCCTC 2280
CAACCCAAC 2289

(5) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3248 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GTCCCTTTCC CTGTCTCCAA GGATTAAAAA AGAATCTCTG ACTGCTGGTT ATCGGCAGGG 60
GTAAAACAAC TGGGTGCAGA ATGGGCACCA ACTAAGCAAA ATCACACACT GGTTGTGAAC 120
AGCTGGAACA ACTGGGTTGG AACATTGCAG CCCAGGAAAG TGATCTTTAA AGACAAGGTG 180
GTCTCTTCAG TGTAATCGAA AGTAATCACT GGGGGATAAT GTTAGCTGTG CCCATGGGAA 240
TGAATGGCCT CCTATCCCAA CTGTACGAGA CAAACACCAT GCCAGATTTG CTCACTGTTG 300
ATGACACTGG GCTGTCTTGT ATCCTAATGA TAAAACAGCT AACTGATAAA GGAGCTTATT 360
TCTCAGGCAG CAGATCTGAG AGATACCTGG GCTTTGAACT GCATGAAAGA GAGAGAAAGT 420
ATCCTAGATA ATGGCTCATG GACTTGAGTG GGCCACTGGA GGTCACCCTT GCTTGCTGCA 480
TATTGGATTA CAGAAAAGAA TTATTTGATT GTTTTGTAGT GTCCTAGCAG CAGGTGATGG 540
TGACAGCCCT TTGAGAGGGT TCAGCCTCAA GGGAGTCCTC CAGGGGGACT CAGAGGGTCC 600
GAAATGTGTC ATTGGCCAAG GTGGGTGGCA ATGCTGTGAC ATTGACCGAG TACATATCAG 660
CCCAGTCCGG GAAGGTGCCC AGCTGGAAGA TGGTCTGTGC CCAGGTATTC ATAGCCAAAC 720
TGAGCAGCAG GACACCCATC AGGTTCATCA GGAGGCCTGT CCGCACCATG TCTTTGACCA 780
GCAAGTGTCC AGAGGCGAAG GCGATGGAGT TGGGGGGCGT TGAGACCGGG AGCATGAAGG 840
CAAAGGAGCA GCCGACTGTG CCCGGAATCA TCAGATACAG GGGGTGCACT CTCAGGCGGA 900
TGGCCAGCTC TGCCAGGACC GGCAGGAAGA TGATGATGGT CGCCGTGTTG CTGGCAAACT 960
CAGTGAAGAA GGCGATGACC ACAGTGATGA GCAGCACAGC CAGGGCGGGG GGCACATTCT 1020
CCAGGGGGTG CAGCTGCCCA CCAATCCATA CAGACAGCCC CGATTCCTCA CAGCCTTTGG 1080
CCATGGCGAA GCCCCCTCCC AGGAGAAGGA TGATGTTCCA AGGCACTGTC TCCTGGGCCT 1140
TCTTCCAGGT CAGCAAGGGC TCTGTCTCTG TGTTGGGAGC TTTGAAGTCA AACCACCACT 1200
TGAGAGAGGG CCTTTGGGAC GGGAAGAAGA ACAAGATGGT GACAATAGCC ACGCCGGTGA 1260
CAGCATCAGA AAGAAACCCA GGATTGAAGA GGCTGGCCCA GCCAGGGATG AACTTCGGGT 1320
CCCGGGTGAA GAGGAGGATG GCAAACATGC AGAAAAGGAT GAAAACAGCC TGTTCGGCAA 1380

ACTTGATGGG CCCCAGGTTC TGGTATTCTT CCCGAATTAC AGCTCGAGCC CTATCTTCTG 1440
CATTGGTTCT TATCTCAGAT TTATTCTTCC TCCAGCCCCT GAAGCTCAGT CCCCCGTACA 1500
GGAAGGAGAT CCAGAGCCAG CCTGCCAACA GGAACAACAG CATAAGAGGG AAGGCGAAAA 1560
TGAACCAGGA GCCGAAATTC ACCACGTCAC ACTGCGGAAA GAAACTCTTG AGCTGGCCAA 1620
GCAGGATGAG GTTAGGGGCT GTGCCCGTGA GTGTGGCTGT GCCCCCAATA CTGGACATGG 1680
ACATGGCAAA TTACAGTGAA GTTTTGGACC CAACTTACAC AACTTTGGAG TTTGAAACTA 1740
TGCAGATTCT ATATAATTCA AGTGATAGTT CTGCCCCAGA GACAAGTATG AATACCACAG 1800
ACAACGGTGT CAACTGTCTG TGTGCTATCT GTGGGGACAG AGCAACAGGA AAACACTATG 1860
GGGCATCCAC CTGTGATGGG TGCAAGGGTT TCTTCAGACG CAGCATTCGT AAGAGTCACA 1920
TTTATTCTTG CAGGTTCAGT CGGCAATGTG TTGTTGACAA GGACAAAAGG AATCAATGTA 1980
GATATTGTCG ATTAAGAAAG TGTTTTAGAG CGGGAATGAA AAAAGAAGCT GTACAAAATG 2040
AACGTGACAG AATAAGCACC AGAAGAAGCA CATTTGATGG CAGCAACATC CCCTCCATTA 2100
ACACACTGGC ACAAGCTGAA GTTCGGTCTC GCCAGATCTC AGTCTCAAGC CCTGGGTCAA 2160
GCACTGACAT AAACGTTAAG AAAATTGCAA GTATTGGTGA TGTCTGTGAA TCTATGAAAC 2220
AGCAGCTCTT AGTCTTGGTG GAATGGGCTA AATATATTCC TGCCTTCTGT GAATTACCAT 2280
TGGATGATCA GGTGGCACTG TTGAGAGCTC ACGCAGGGGA GCACTTACTG CTTGGAGCTA 2340
CAAAGAGATC CATGATGTAT AAAGATATTT TGCTTTTGGG AAACAACTAT GTTATTCACC 2400
GCAACAGCTG TGAAGTTGAG ATTAGCCGTG TGGCCAATCG TGTTCTAGAT GAGCTGGTTA 2460
GACCATTTCA AGAAATCCAG ATTGATGACA ATGAGTATGC TTGTTTAAAG GCAATTGTAT 2520
TTTTTGATCC AGATGCAAAA GGGCTAAGCG ATCCAGTAAA AATTAAGAAC ATGAGGTTCC 2580
AAGTGCAGAT CGGTTTGGAG GACTACATCA ATGATCGGCA GTATGACTCC CGGGGGAGGT 2640
TTGGAGAGTT GCTTCTGCTC CTGCCCACAC TGCAGAGCAT CACGTGGCAA ATGATTGAGC 2700
AAATACAGTT TGTTAAACTT TTTGGGATGG TTAAAATTGA CAATCTACTT CAGGAAATGC 2760
TATTAGGTGG GGCTTCCAAT GATGGCAGTC ATCTCCATCA TCCAATGCAT CCACATTTGT 2820
CTCAAGACCC ATTAACTGGA CAAACTATAC TTTTAGGTCC CATGTCAACA CTGGTTCATG 2880
CAGACCAGAT CTCAACTCCT GAAACCCCAC TCCCTTCCCC ACCACAAGGC TCTGGGCAAG 2940
AACAGTACAA AATAGCTGCA AACCAAGCAT CAGTCATTTC ACACCAGCAT CTCTCCAAAC 3000
AAAAGCAATT GTGAAAATGT GTTTACTTCA GAACGGCACT ACATAAATGT GAAAAGTTGT 3060
TGATCTTGAA ATATCTCAAG ATAGCACTTT TGGCAAACTC TTAGCCAAGG CTTCTTCATT 3120
GGTGCTGTTA TAAGATGGTA TCCTATTTTC TTGTTTATAC GTTCATTCTA TTTGTTATTG 3180
CTACTATGTG AAACTTTCAC ATGCAACCAA TGTATATCTG AGTTTGAAGG ATGTTTATAT 3240
AGGGTAGG 3248

CLAIMS

1. A method for determining if an animal is at risk for diabetes, comprising:

providing an animal; and

evaluating an aspect of HNF4 metabolism or structure in said animal, an abnormality in said aspect of HNF4 metabolism or structure being diagnostic of being at risk for diabetes.

2. The method of claim 1 wherein said diabetes is non-insulin dependent diabetes (NIDDM).

3. The method of claim 1 wherein said diabetes is a common form of non-insulin-dependent diabetes (NIDDM).

4. The method of claim 1 wherein said diabetes is maturity-onset diabetes of the young (MODY).

5. The method of claim 4 wherein said diabetes is MODY 1.

6. The method of claim 1 wherein said diabetes is an insulin secretion-related disorder.

7. The method of claim 1 wherein said animal is a prenatal animal.

8. A method for evaluating an agent for use in treating diabetes, comprising:

providing a test cell, cell-free system or animal having a non-wild type pattern of HNF4 metabolism;

providing an agent;

administering said agent to said test cell, cell-free system or animal in a therapeutically effective amount; and

evaluating the effect of said agent on an aspect of HNF4 metabolism, a change in said aspect of HNF4 metabolism being indicative of the usefulness of said agent in treating diabetes.

9. The method of claim 8 wherein said agent comprises HNF4 polypeptide or a biologically active fragment or analog thereof.

10. The method of claim 8 wherein said agent comprises a nucleic acid encoding HNF4 polypeptide or a biologically active fragment thereof.

11. The method of claim 8 wherein said agent is selected from the group consisting of a nucleic acid comprising a nucleotide sequence as set forth in Fig. 1 (SEQ. ID NO:1), Fig. 2 (SEQ. ID NO:2) and Fig. 3 (SEQ. ID NO:3).

12. The method of claim 8 wherein said agent is a nucleic acid comprising a nucleotide sequence as set forth in Fig. 4 (SEQ. ID NO:4).

13. The method of claim 8 wherein said agent comprises a nucleic acid encoding an HNF4 regulatory sequence or a biologically active fragment thereof.

14. The method of claim 8 wherein said agent is selected from the group consisting of a binding molecule for HNF4 polypeptide and a binding molecule for HNF4 nucleic acid.

15. The method of claim 8 wherein said agent is an antisense nucleic acid.

16. The method of claim 8 wherein said agent is selected from the group consisting of a mimetic of HNF4 and a mimetic of a binding molecule of HNF4.

17. The method of claim 8 wherein said agent is an antibody.

18. The method of claim 8 wherein said agent is a natural ligand for HNF4.

19. The method of claim 8 wherein said agent is an artificial ligand for HNF4.

20. The method of claim 8 wherein said agent is selected from the group consisting of an antagonist, an agonist and a super agonist.

21. The method of claim 8 wherein said agent is administered to a member selected from the group consisting of a transgenic cell and a transgenic animal.

22. The agent identified in claim 8.

23. The method of claim 8 wherein said agent is administered to said test cell or cell-free system in vitro, and if said change in said aspect of said HNF metabolism occurs, then further administering said agent to a test animal in a therapeutically effective amount and evaluating the in vivo effect of said agent on an aspect of HNF4 metabolism.

24. A method for evaluating an agent for the ability to alter the binding of HNF4 polypeptide to a binding molecule, comprising:

providing an agent;

providing HNF4 polypeptide;

providing a binding molecule;

combining said agent, said HNF4 polypeptide and said binding molecule; and

detecting the formation of a complex comprising said HNF4 polypeptide and said binding molecule, an alteration in the formation of said complex in the presence of said agent as compared to in the absence of said agent being indicative of said agent altering the binding of said HNF4 polypeptide to said binding molecule.

25. The method of claim 24 wherein the altering of the binding of said HNF4 polypeptide to said binding molecule is inhibiting the binding.

26. The method of claim 24 wherein the altering of the binding of said HNF4 polypeptide to said binding molecule is promoting the binding.

27. The agent identified in claim 24.

28. A method for evaluating an agent for the ability to bind to HNF4 polypeptide, comprising:

providing an agent;

providing an HNF4 polypeptide;

contacting said agent with said HNF4 polypeptide; and

evaluating the ability of said agent to bind to said HNF4 polypeptide.

29. The agent identified in claim 28.

30.A method for evaluating an agent for the ability to bind to a nucleic acid encoding an HNF4 regulatory sequence, comprising:

providing an agent;

providing a nucleic acid encoding an HNF4 regulatory sequence;

contacting said agent with said nucleic acid; and

evaluating the ability of said agent to bind to said nucleic acid.

31. The agent identified in claim 30.

32. A method for treating diabetes in an animal, comprising:

providing an animal in need of treatment for diabetes;

providing an agent capable of altering an aspect of HNF4 structure or metabolism;

administering said agent to said animal in a therapeutically effective amount such that treatment of said diabetes occurs.

33. A method for treating an animal at risk for diabetes, comprising:

providing an animal at risk for diabetes;

providing an agent capable of altering an aspect of HNF4 structure or metabolism; and

administering said agent to said animal in a therapeutically effective amount such that treatment of said animal occurs.

34. A method for treating a cell having an abnormality in structure or metabolism of HNF4, comprising:

providing a cell having an abnormality in structure or metabolism of HNF4;

providing an agent capable of altering an aspect of HNF4 structure or metabolism; and

administering said agent to said cell in a therapeutically effective amount such that treatment of said cell occurs.

35. The method of claim 34 wherein said cell is obtained from a cell or tissue culture.

36. The method of claim 34 wherein said cell is obtained from an embryo fibroblast.

37. The method of claim 34 wherein said cell is part of an animal.

38. The method of claim 37 wherein said animal is a non-human transgenic animal.

39. A pharmaceutical composition for treating diabetes in an animal, comprising:

a therapeutically effective amount of an agent, said agent being capable of altering an aspect of HNF4 metabolism or structure in said animal so as to result in treatment of said diabetes; and

a pharmaceutically acceptable carrier.

40. A vaccine composition for treating diabetes in an animal, comprising:

a therapeutically effective amount of an agent, said agent being capable of altering an aspect of HNF4 metabolism or structure in said animal so as to result in treatment of said diabetes; and

a pharmaceutically acceptable carrier.

41. A method of making a fragment or analog of HNF4 polypeptide, said fragment or analog having a transcription activating activity of a naturally occurring HNF4 polypeptide for a gene involved in insulin metabolism, comprising:

providing an HNF4 polypeptide;

altering the sequence of said HNF4 polypeptide; and

testing said altered HNF4 polypeptide for said transcription activating activity.

42. A method of making an HNF4 polypeptide or fragment or analog thereof having a non-wild type activity as a transcription activator for a gene involved in insulin metabolism, comprising:

providing an HNF4 polypeptide;

altering the sequence of said HNF4 polypeptide; and

testing said altered HNF4 polypeptide for said non-wild type activity.

43. The method of claim 42 wherein said altered HNF4 polypeptide is selected from the group consisting of an antagonist, an agonist and a super agonist.

FIG. 1

```
GGGGCGGAGG GCGGGGGCTT CGGGGTGGGC GCCCAGGGTA GGGCAGGTGG CCGCCGCGTG      60
GAGGCAGGGA GAATGCACTC TCCAAAACCC TCCGTCGACA TGGACATGGC CGACTACAGT     120
GCTGCACTGG ACCCAGCCTA CACCACCCTG GAATTTGAGA ATGTGCAGGT GTTGACGATG     180
GGCAATGACA CGTCCCCATC AGAAGGCACC AACCTCAACG CGCCCAACAG CCTGGGTGTC     240
AGCGCCCTGT GTGCCATCTG CGGGGACCGG GCCACGGGCA AACACTACGG TGCCTCGAGC     300
TGTGACGGCT GCAAGGGCTT CTTCCGGAGG AGCGTGCGGA AGAACCACAT GTACTCCTGC     360
AGATTTAGCC GGCAGTGCGT GGTGGACAAG GACAAGAGGA ACCAGTGCCG CTACTGCAGG     420
CTCAAGAAAT GCTTCCGGGC TGGCATGAAG AAGGAAGCCG TCCAGAATGA GCGGGACCGG     480
ATCAGCACTC GAAGGTCAAG CTATGAGGAC AGCAGCCTGC CCTCCATCAA TGCGCTCCTG     540
CAGGCGGAGG TCCTGTCCCG ACAGATCACC TCCCCCGTCT CCGGGATCAA CGGCGACATT     600
CGGGCGAAGA AGATTGCCAG CATCGCAGAT GTGTGTGAGT CCATGAAGGA GCAGCTGCTG     660
GTTCTCGTTG AGTGGGCCAA GTACATCCCA GCTTTCTGCG AGCTCCCCCT GGACGACCAG     720
GTGGCCCTGC TCAGAGCCCA TGCTGGCGAG CACCTGCTGC TCGGAGCCAC CAAGAGATCC     780
ATGGTGTTCA AGGACGTGCT GCTCCTAGGC AATGACTACA TTGTCCCTCG GCACTGCCCG     840
GAGCTGGCGG AGATGAGCCG GGTGTCCATA CGCATCCTTG ACGAGCTGGT GCTGCCCTTC     900
CAGGAGCTGC AGATCGATGA CAATGAGTAT GCCTACCTCA AAGCCATCAT CTTCTTTGAC     960
CCAGATGCCA AGGGGCTGAG CGATCCAGGG AAGATCAAGC GGCTGCGTTC CCAGGTGCAG    1020
GTGAGCTTGG AGGACTACAT CAACGACCGC CAGTATGACT CGCGTGGCCG CTTTGGAGAG    1080
CTGCTGCTGC TGCTGCCCAC CTTGCAGAGC ATCACCTGGC AGATGATCGA GCAGATCCAG    1140
TTCATCAAGC TCTTCGGCAT GGCCAAGATT GACAACCTGT TGCAGGAGAT GCTGCTGGGA    1200
GGGTCCCCCA GCGATGCACC CCATGCCCAC CACCCCCTGC ACCCTCACCT GATGCAGGAA    1260
CATATGGGAA CCAACGTCAT CGTTGCCAAC ACAATGCCCA CTCACCTCAG CAACGGACAG    1320
ATGTCCACCC CTGAGACCCC ACAGCCCTCA CCGCCAGGTG GCTCAGGGTC TGAGTCCTAT    1380
AAGCTCCTGC CGGGAGCCGT CGCCACAATC GTCAAGCCCC TCTCTGCCAT CCCCCAGCCG    1440
ACCATCACCA AGCAGGAAGT TATCTAGCAA GCCGCTGGGG CTTGGGGGCT CCACTGGCTC    1500
CCCCCAGTCC CCCTAAGAGA GCACCTGGTG ATACCGTGGT CACGGCAAAG GAAGACGTGA    1560
TGCCAGGACC AGTCCCAGAG CAGGAATGGG AAGGATGAAG GGCCCGAGAA CATGGCCTAA    1620
GGGCACATCC CACTGCCACC CTTGACGCCC CTGCTCTGGA TAACAAGACT TTGACTTGGG    1680
GAGACCTCTA CTGCCTTGGA CAACTTTTCT CATGTTGAAG CCACTGCCTT CACCTTCACC    1740
TTCATCCATG TCCAACCCCC GACTTCATCC CAAAGGACAG CCGCCTGGAG ATGACTTGAG    1800
GCCTTACTTA AACCCAGCTC CCTGTGCCCT AGCCTGGTGC TTCTCCTCTC CTAGCCCCTG    1860
TCATGGTGTC CCAGACAGAG CCCTGTGAGG CTGGGTCCAA TTGTGGCACT TGGGGCACCT    1920
TGCTCCTCCT TCTGCTGCTG CCCCCACCTC TGCTGTCACC TTGCTCAGCC ATCCCGTCTT    1980
CTCCAACACC ACCTCTCCAG AGGCCAAGGA GGCCTTGGAA ACGATTCCCC CAGTCATTCT    2040
GGGAATACTG TTGTAAGACC TGACTGGGCC AGGCACCAGG CAGGGTCTAG AAGGCTGTGG    2100
TGAGGGAAGA CGCCTTTCTC CTCCAACCCA ACCTCATCCT CCTTCTTCAG GGACTTGGGT    2160
GGTACTTGGG TGAGGATCCC TAAGGCCTTC AACCCGAGAA AACAAACCCA GGTTGGCGAC    2220
TGCAACAGGA ACTTGGAGTG GAGAGGAAAA GCATCAGAAA GAGGCAGACC ATCCACCAGG    2280
CCTTTGAG                                                            2288
```

FIG. 2

```
GGCTTCGGGG TGGGCGCCCA GGGTAGGGCA GGTGGCCGCC GCGTGGAGGC AGGGAGAATG      60
CACTCTCCAA AACCCTCCGT CGACATGGAC ATGGCCGACT ACAGTGCTGC ACTGGACCCA     120
GCCTACACCA CCCTGGAATT TGACAATGTG CAGGTGTTGA CGATGGGCAA TGACACGTCC     180
CCATCAGAAG GCACCAACCT CAACGCGCCC AACAGCCTGG GTGTCAGCGC CCTGTGTGCC     240
ATCTGCGGGG ACCGGGCCAC GGGCAAACAC TACGGTGCCT CGAGCTGTGA CGGCTGCAAG     300
GGCTTCTTCC GGAGGAGCGT GCGGAAGAAC CACATGTACT CCTGCAGATT TAGCCGGCAG     360
TGCGTGGTGG ACAAGGACAA GAGGAACCAG TGCCGCTACT GCAGGCTCAA GAAATGCTTC     420
CGGGCTGGCA TGAAGAAGGA AGCCGTCCAG AATGAGCGGG ACCGGATCAG CACTCGAAGG     480
TCAAGCTATG AGGACAGCAG CCTGCCCTCC ATCAATGCGC TCCTGCAGGC GGAGGTCCTG     540
TCCCGACAGA TCACCTCCCC CGTCTCCGGG ATCAACGGCG ACATTCGGGC GAAGAAGATT     600
GCCAGCATCG CAGATGTGTG TGAGTCCATG AAGGAGCAGC TGCTGGTTCT CGTTGAGTGG     660
GCCAAGTACA TCCCAGCTTT CTGCGAGCTC CCCCTGGACG ACCAGGTGGC CCTGCTCAGA     720
GCCCATGCTG GCGAGCACCT GCTGCTCGGA GCCACCAAGA GATCCATGGT GTTCAAGGAC     780
GTGCTGCTCC TAGGCAATGA CTACATTGTC CCTCGGCACT GCCCGGAGCT GGCGGAGATG     840
AGCCGGGTGT CCATACGCAT CCTTGACGAG CTGGTGCTGC CCTTCCAGGA GCTGCAGATC     900
GATGACAATG AGTATGCCTA CCTCAAAGCC ATCATCTTCT TTGACCCAGA TGCCAAGGGG     960
CTGAGCGATC CAGGGAAGAT CAAGCGGCTG CGTTCCCAGG TGCAGGTGAG CTTGGAGGAC    1020
TACATCAACG ACCGCCAGTA TGACTCGCGT GGCCGCTTTG GAGAGCTGCT GCTGCTGCTG    1080
CCCACCTTGC AGAGCATCAC CTGGCAGATG ATCGAGCAGA TCCAGTTCAT CAAGCTCTTC    1140
GGCATGGCCA AGATTGACAA CCTGTTGCAG GAGATGCTGC TGGGAGGGTC CCCCAGCGAT    1200
GCACCCCATG CCCACCACCC CCTGCACCCT CACCTGATGC AGGAACATAT GGGAACCAAC    1260
GTCATCGTTG CCAACACAAT GCCCACTCAC CTCAGCAACG GACAGATGTG TGAGTGGCCC    1320
CGACCCAGGG GACAGGCAGC CACCCCTGAG ACCCCACAGC CCTCACCGCC AGGTGGCTCA    1380
GGGTCTGAGT CCTATAAGCT CCTGCCGGGA GCCGTCGCCA CAATCGTCAA GCCCCTCTCT    1440
GCCATCCCCC AGCCGACCAT CACCAAGCAG GAAGTTATCT AGCAAGCCGC TGGGGCTTGG    1500
GGGCTCCACT GGCTCCCCCC AGTCCCCCTA AGAGAGCACC TGGTGATACC GTGGTCACGG    1560
CAAAGGAAGA CGTGATGCCA GGACCAGTCC CAGAGCAGGA ATGGGAAGGA TGAAGGGCCC    1620
GAGAACATGG CCTAA                                                     1635
```

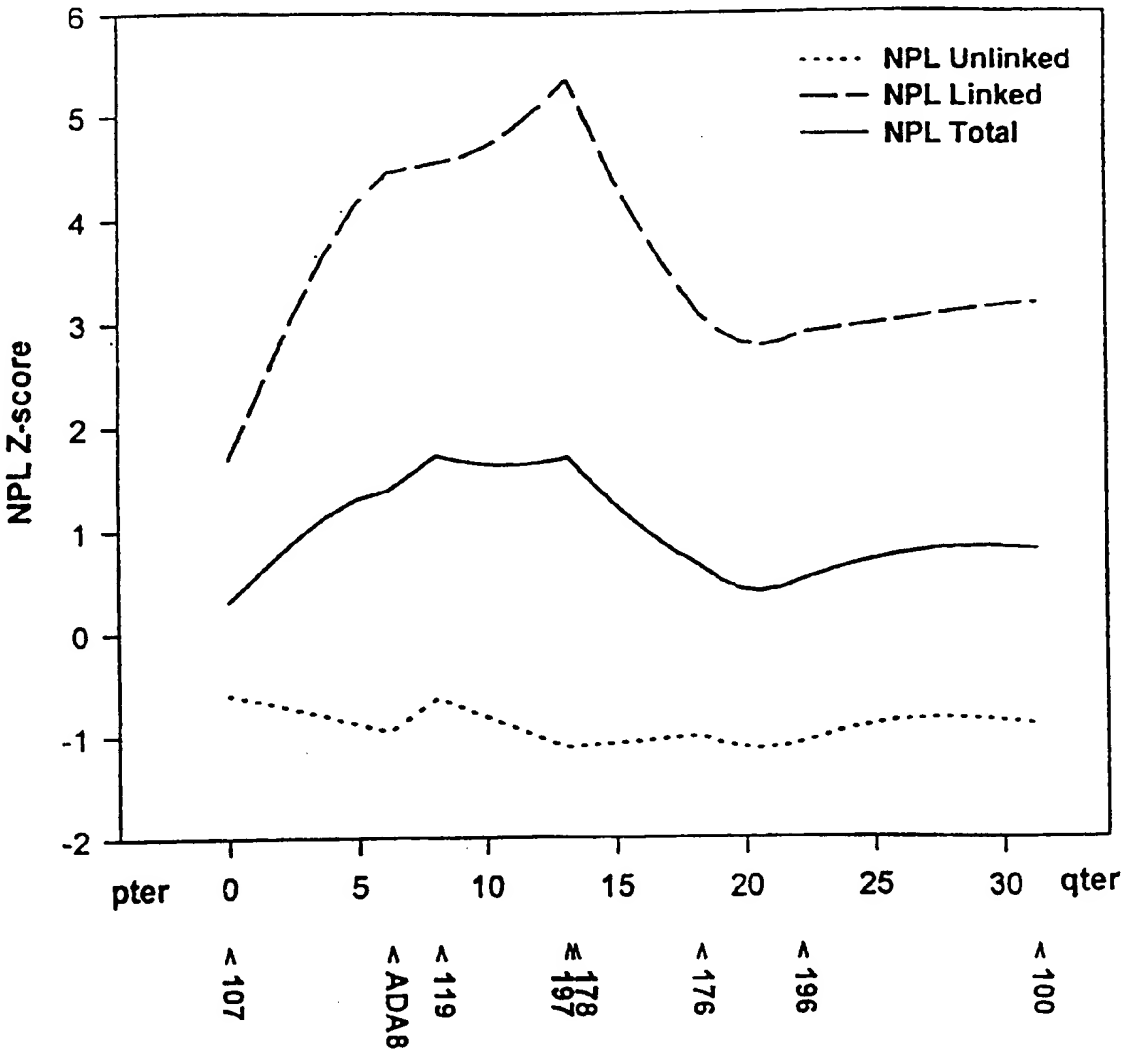
FIG. 3

```
GCAGAGAGGG CACTGGGAGG AGGCAGTGGG AGGGCGGAGG GCGGGGGCCT TCGGGGTGGG      60
CGCCCAGGGT AGGGCAGGTG GCCGCGGCGT GGAGGCAGGG AGAATGCGAC TCTCCAAAAC     120
CCTCGTCGAC ATGGACATGG CCGACTACAG TGCTGCACTG GACCCAGCCT ACACCACCCT     180
GGAATTTGAG AATGTGCAGG TGTTGACGAT GGGCAATGAT TTGTTGCCGC TGCGTCTCGC     240
CAGATTGAGG CATCCCCTCC GACATCACTG GAGCATATCT GGAGGGGTGG ACAGTTCTCC     300
ACAGGGAGAC ACGTCCCCAT CAGAAGGCAC CAACCTCAAC GCGCCCAACA GCCTGGGTGT     360
CAGCGCCCTG TGTGCCATCT GCGGGGACCG GGCCACGGGC AAACACTACG GTGCCTCGAG     420
CTGTGACGGC TGCAAGGGCT TCTTCCGGAG GAGCGTGCGG AAGAACCACA TGTACTCCTG     480
CAGATTTAGC CGGCAGTGCG TGGTGGACAA AGACAAGAGG AACCAGTGCC GCTACTGCAG     540
GCTCAAGAAA TGCTTCCGGG CTGGCATGAA GAAGGAAGCC GTCCAGAATG AGCGGGACCG     600
GATCAGCACT CGAAGGTCAA GCTATGAGGA CAGCAGCCTG CCCTCCATCA ATGCGCTCCT     660
GCAGGCGGAG GTCCTGTCCC GACAGATCAC CTCCCCCGTC TCCGGGATCA ACGGCGACAT     720
TCGGGCGAAG AAGATTGCCA GCATCGCAGA TGTGTGTGAG TCCATGAAGG AGCAGCTGCT     780
GGTTCTCGTT GAGTGGGCCA AGTACATCCC AGCTTTCTGC GAGCTCCCCC TGGACGACCA     840
GGTGGCCCTG CTCAGAGCCC ATGCTGGCGA GCACCTGCTG CTCGGAGCCA CCAAGAGATC     900
CATGGTGTTC AAGGACGTGC TGCTCCTAGG CAATGACTAC ATTGTCCCTC GGCACTGCCC     960
GGAGCTGGCG GAGATGAGCC GGGTGTCCAT ACGCATCCTT GACGAGCTGG TGCTGCCCTT    1020
CCAGGAGCTG CAGATCGATG ACAATGAGTA TGCCTACCTC AAAGCCATCA TCTTCTTTGA    1080
CCCAGATGCC AAGGGGCTGA GCGATCCAGG GAAGATCAAG CGGCTGCGTT CCCAGGTGCA    1140
GGTGAGCTTG GAGGACTACA TCAACGACCG CCAGTATGAC TCGCGTGGCC GCTTTGGAGA    1200
GCTGCTGCTG CTGCTGCCCA CCTTGCAGAG CATCACGTGG CAGATGATCG AGCAGATCCA    1260
GTTCATCAAG CTCTTCGGCA TGGCCAAGAT TGACAACCTG TTGCAGGAGA TGCTGCTGGG    1320
AGGGTCCCCC AGCGATGCAC CCCATGCCCA CCACCCCCTG CACCCTCACC TGATGCAGGA    1380
ACATATGGGA ACCAACGTCA TCGTTGCCAA CACAATGCCC ACTCACCTCA GCAACGGACA    1440
GATGTGTGAG TGGCCCCGAC CCAGGGGACA GGCAGCCACC CCTGAGACCC CACAGCCCTC    1500
ACCGCCAGGT GGCTCAGGGT CTGAGCCCTA TAAGCTCCTG CCGGGAGCCG TCGCCACAAT    1560
CGTCAAGCCC CTCTCTGCCA TCCCCCAGCC GACCATCACC AAGCAGGAAG TTATCTAGCA    1620
AGCCGCTGGG GCTTGGGGGC TCCACTGGCT CCCCCCAGCC CCCTAAGAGA GCACCTGGTG    1680
ATCACGTGGT CACGGCAAAG GAAGACGTGA TGCCAGGACC AGTCCCAGAG CAGGAATGGG    1740
AAGGATGAAG GGCCCGAGAA CATGGCCTAA GGCACATCCC ACTGCACCCT GACGCCCTGC    1800
TCTGATAACA AGACTTTGAC TTGGGGAGAC CCTCTACTGC CTTGGACAAC TTTCTCATGT    1860
TGAAGCCACT GCCTTCACCT TCACCTTCAT CCATGTCCAA CCCCCGACTT CATCCCAAAG    1920
GACAGCCGCC TGGAGATGAC TTGAGCCTTA CTTAAACCCA GCTCCCTTCT TCCCTAGCCT    1980
GGTGCTTCTC CTCTCCTAGC CCCGGTCATG GTGTCCAGAC AGAGCCCTGT GAGGCTGGGT    2040
CCAATTGTGG CACTTGGGGC ACCTTGCTCC TCCTTCTGCT GCTGCCCCCA CCTCTGCTGC    2100
CTCCCTCTGC TGTCACCTTG CTCAGCCATC CCGTCTTCTC CAACACCACC TCTACAGAGG    2160
CCAAGGAGGC CTTGGAAACG ATTCCCCCAG TCATTCTGGG AACATGTTGT AAGCACTGAC    2220
TGGGACCAGG CACCAGGCAG GGTCTAGAAG GCTGTGGTGA GGGAAGACGC CTTTCTCCTC    2280
CAACCCAAC                                                            2289
```

FIG. 4

```
GTCCCTTTCC CTGTCTCCAA GGATTAAAAA AGAATCTCTG ACTGCTGGTT ATCGGCAGGG      60
GTAAAACAAC TGGGTGCAGA ATGGGCACCA ACTAAGCAAA ATCACACACT GGTTGTGAAC     120
AGCTGGAACA ACTGGGTTGG AACATTGCAG CCCAGGAAAG TGATCTTTAA AGACAAGGTG     180
GTCTCTTCAG TGTAATCGAA AGTAATCACT GGGGGATAAT GTTAGCTGTG CCCATGGGAA     240
TGAATGGCCT CCTATCCCAA CTGTACGAGA CAAACACCAT GCCAGATTTG CTCACTGTTG     300
ATGACACTGG GCTGTCTTGT ATCCTAATGA TAAAACAGCT AACTGATAAA GGAGCTTATT     360
TCTCAGGCAG CAGATCTGAG AGATACCTGG GCTTTGAACT GCATGAAAGA GAGAGAAAGT     420
ATCCTAGATA ATGGCTCATG GACTTGAGTG GGCCACTGGA GGTCACCCTT GCTTGCTGCA     480
TATTGGATTA CAGAAAAGAA TTATTTGATT GTTTTGTAGT GTCCTAGCAG CAGGTGATGG     540
TGACAGCCCT TTGAGAGGGT TCAGCCTCAA GGGAGTCCTC CAGGGGGACT CAGAGGGTCC     600
GAAATGTGTC ATTGGCCAAG GTGGGTGGCA ATGCTGTGAC ATTGACCGAG TACATATCAG     660
CCCAGTCCGG GAAGGTGCCC AGCTGGAAGA TGGTCTGTGC CCAGGTATTC ATAGCCAAAC     720
TGAGCAGCAG GACACCCATC AGGTTCATCA GGAGGCCTGT CCGCACCATG TCTTTGACCA     780
GCAAGTGTCC AGAGGCGAAG GCGATGGAGT TGGGGGGCGT TGAGACCGGG AGCATGAAGG     840
CAAAGGAGCA GCCGACTGTG CCCGGAATCA TCAGATACAG GGGGTGCACT CTCAGGCGGA     900
TGGCCAGCTC TGCCAGGACC GGCAGGAAGA TGATGATGGT CGCCGTGTTG CTGGCAAACT     960
CAGTGAAGAA GGCGATGACC ACAGTGATGA GCAGCACAGC CAGGGCGGGG GGCACATTCT    1020
CCAGGGGGTG CAGCTGCCCA CCAATCCATA CAGACAGCCC CGATTCCTCA CAGCCTTTGG    1080
CCATGGCGAA GCCCCCTCCC AGGAGAAGGA TGATGTTCCA AGGCACTGTC TCCTGGGCCT    1140
TCTTCCAGGT CAGCAAGGGC TCTGTCTCTG TGTTGGGAGC TTTGAAGTCA AACCACCACT    1200
TGAGAGAGGG CCTTTGGGAC GGGAAGAAGA ACAAGATGGT GACAATAGCC ACGCCGGTGA    1260
CAGCATCAGA AAGAAACCCA GGATTGAAGA GGCTGGCCCA GCCAGGGATG AACTTCGGGT    1320
CCCGGGTGAA GAGGAGGATG GCAAACATGC AGAAAAGGAT GAAAACAGCC TGTTCGGCAA    1380
ACTTGATGGG CCCCAGGTTC TGGTATTCTT CCCGAATTAC AGCTCGAGCC CTATCTTCTG    1440
CATTGGTTCT TATCTCAGAT TTATTCTTCC TCCAGCCCCT GAAGCTCAGT CCCCCGTACA    1500
GGAAGGAGAT CCAGAGCCAG CCTGCCAACA GGAACAACAG CATAAGAGGG AAGGCGAAAA    1560
TGAACCAGGA GCCGAAATTC ACCACGTCAC ACTGCGGAAA GAAACTCTTG AGCTGGCCAA    1620
GCAGGATGAG GTTAGGGGCT GTGCCCGTGA GTGTGGCTGT GCCCCCAATA CTGGACATGG    1680
ACATGGCAAA TTACAGTGAA GTTTTGGACC CAACTTACAC AACTTTGGAG TTTGAAACTA    1740
TGCAGATTCT ATATAATTCA AGTGATAGTT CTGCCCCAGA GACAAGTATG AATACCACAG    1800
ACAACGGTGT CAACTGTCTG TGTGCTATCT GTGGGGACAG AGCAACAGGA AAACACTATG    1860
GGGCATCCAC CTGTGATGGG TGCAAGGGTT TCTTCAGACG CAGCATTCGT AAGAGTCACA    1920
TTTATTCTTG CAGGTTCAGT CGGCAATGTG TTGTTGACAA GGACAAAAGG AATCAATGTA    1980
GATATTGTCG ATTAAGAAAG TGTTTTAGAG CGGGAATGAA AAAAGAAGCT GTACAAAATG    2040
AACGTGACAG AATAAGCACC AGAAGAAGCA CATTTGATGG CAGCAACATC CCCTCCATTA    2100
ACACACTGGC ACAAGCTGAA GTTCGGTCTC GCCAGATCTC AGTCTCAAGC CCTGGGTCAA    2160
GCACTGACAT AAACGTTAAG AAAATTGCAA GTATTGGTGA TGTCTGTGAA TCTATGAAAC    2220
AGCAGCTCTT AGTCTTGGTG GAATGGGCTA AATATATTCC TGCCTTCTGT GAATTACCAT    2280
TGGATGATCA GGTGGCACTG TTGAGAGCTC ACGCAGGGGA GCACTTACTG CTTGGAGCTA    2340
CAAAGAGATC CATGATGTAT AAAGATATTT TGCTTTTGGG AAACAACTAT GTTATTCACC    2400
GCAACAGCTG TGAAGTTGAG ATTAGCCGTG TGGCCAATCG TGTTCTAGAT GAGCTGGTTA    2460
GACCATTTCA AGAAATCCAG ATTGATGACA ATGAGTATGC TTGTTTAAAG GCAATTGTAT    2520
TTTTTGATCC AGATGCAAAA GGGCTAAGCG ATCCAGTAAA AATTAAGAAC ATGAGGTTCC    2580
AAGTGCAGAT CGGTTTGGAG GACTACATCA ATGATCGGCA GTATGACTCC CGGGGGAGGT    2640
TTGGAGAGTT GCTTCTGCTC CTGCCCACAC TGCAGAGCAT CACGTGGCAA ATGATTGAGC    2700
AAATACAGTT TGTTAAACTT TTTGGGATGG TTAAAATTGA CAATCTACTT CAGGAAATGC    2760
TATTAGGTGG GGCTTCCAAT GATGGCAGTC ATCTCCATCA TCCAATGCAT CCACATTTGT    2820
CTCAAGACCC ATTAACTGGA CAAACTATAC TTTTAGGTCC CATGTCAACA CTGGTTCATG    2880
CAGACCAGAT CTCAACTCCT GAAACCCCAC TCCCTTCCCC ACCACAAGGC TCTGGGCAAG    2940
AACAGTACAA AATAGCTGCA AACCAAGCAT CAGTCATTTC ACACCAGCAT CTCTCCAAAC    3000
AAAAGCAATT GTGAAAATGT GTTTACTTCA GAACGGCACT ACATAAATGT GAAAAGTTGT    3060
TGATCTTGAA ATATCTCAAG ATAGCACTTT TGGCAAACTC TTAGCCAAGG CTTCTTCATT    3120
GGTGCTGTTA TAAGATGGTA TCCTATTTTC TTGTTTATAC GTTCATTCTA TTTGTTATTG    3180
CTACTATGTG AAACTTTCAC ATGCAACCAA TGTATATCTG AGTTTGAAGG ATGTTTATAT    3240
AGGGTAGG                                                            3248
```

FIG. 5


NPL Unlinked
NPL Linked
NPL Total
NPL Z-score
6
5
4
3
2
1
0
-1
-2
pter
0
5
10
15
20
25
30
qter
107
ADA8
119
178
197
176
196
100

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/21614

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.
US CL :435/4, 6, 7.1, 68.1, 69.1, 172.1; 514/12; 530/350
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/4, 6, 7.1, 68.1, 69.1, 172.1; 514/12; 530/350

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	STOFFEL et al. A yeast artificial chromosome-based map of the region of chromosome 20 containing the diabetes-susceptibility gene, MODY1, and a myeloid leukemia related gene. Proceedings of the National Academy of Sciences, USA. April 1996, Vol. 93, pages 3937-3941, especially page 3941.	1-40
Y,P	YAMAGATA et al. Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1). Nature. 05 December 1996, Vol. 384, no. 6608, pages 458-460, see entire document.	1-40
A,P	TODD, J.A. Transcribing diabetes. Nature. 05 December 1996, Vol. 384, no. 6608, pages 407-408, see entire article.	1-40

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance, the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search
28 JANUARY 1998

Date of mailing of the international search report
23 FEB 1998

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
KAWAI LAU
Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/21614

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A,P	HOPKINS, K.D. Genes found for maturity-onset diabetes of the young. The Lancet. 07 December 1996, Vol. 348, no. 9041, page 1573, see entire article.	1-40
Y,P	US 5,604,115 A (SLADEK et al.) 18 February 1997, col. 16, lines 20-24 and col. 16, line 66 to col. 17, line 5.	41-43

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/21614

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

C12Q 1/68; C12P 21/06; C12N 15/12; A61K 38/17 ; C07K 14/47, 14/705

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS, STN-Sciscarch, CAplus, Biosis, Medline, Embase, Toxlit, Cancerlit, Lifesciences, Toxline, USpatfull, CJACS, Genbank, CABA, Biobusiness, Dissabstr, Embal, DGene, Promt, Aquasci, Biotechds, Confsci, JICST-Eplus, PHIN, Europatfull, PATOSWO
search terms: HNF4, diabetes, insulin, NIDDM, MODY, fragment, analog

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s) 1-7, drawn to methods for determining if an animal is at risk for diabetes comprising evaluating an aspect of HNF4 metabolism or structure.
Group II, claim(s) 8-23, 32, 33, 39 and 40, drawn to the combination category of methods for evaluating agents for use in treating diabetes, such agents *per se* as well as pharmaceutical and vaccine compositions comprising them, and methods of treating diabetes or the risk thereof comprising administration of the agents.
Group III, claim(s) 24-29, drawn to the combination category of methods for evaluating agents for the ability to alter HNF4 binding to other molecules as well as the ability to bind HNF4 directly and such agents *per se*.
Group IV, claim(s) 30 and 31, drawn to the combination category of methods for evaluating agents for the ability to bind to HNF4 regulatory sequences and such agents *per se*.
Group V, claim(s) 34-38, drawn to methods of treating cells having an abnormality in HNF4 structure or metabolism.
Group VI, claim(s) 41-43, drawn to methods of making HNF4 fragments or analogs.

The inventions listed as Groups do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: as Stoffel et al. ("A yeast artificial chromosome-based map of the region of chromosome 20 containing the diabetes-susceptibility gene, MODY1, and a myeloid leukemia related gene" Proceedings of the National Academy of Sciences, USA. (April 1996), Vol. 93, pages 3937-3941) teach identification of HNF4 as a candidate gene located in the interval as containing MODY1 (a diabetes susceptibility gene), the concept of HNF4 involvement in diabetes does not constitute a special technical feature, defined as a contribution over the prior art.

Thus the special technical feature of Group I is the method of determining risk of diabetes, which is not shared by Groups II-VI.
The special technical feature of Group II is the method of identifying agents for treating diabetes, which is not shared by Groups I and III-VI.
The special technical feature of Group III is the method of identifying agents to alter HNF4 binding and to bind HNF4, which is not shared by Groups I, II and IV-VI.
The special technical feature of Group IV is the method of identifying agents that bind HNF4 regulatory sequences, which is not shared by Groups I-III, V and VI.
The special technical feature of Group V is the method of treating cells, which is not shared by Groups I-IV and VI.
The special technical feature of Group VI is the method of making HNF4 fragments and analogs, which is not shared by Groups I-IV.

Since Groups I-VI do not share a corresponding special technical feature, they are not so linked as to form a single inventive concept under PCT Rule 13.1.